Universidade de Aveiro
2013
Departamento de Electrónica, Telecomunicações e Informática

Pedro Miguel Mimoso Direito

Decomposição da Latência por segmento numa Rede Móvel de Banda Larga

Universidade de Aveiro
2013
Departamento de Electrónica, Telecomunicações e Informática

Pedro Miguel Mimoso Direito

Decomposição da Latência por segmento numa Rede Móvel de Banda Larga

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica da Prof. Dra. Susana Sargento, Professora auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

Dedico este trabalho à Sofia, Miguel e Vicente.

o júri

Presidente

Professor Doutor Rui Luis Andrade Aguiar

Professor Associado com Agregação, Universidade de Aveiro

Vogais

Professora Doutora Susana Isabel Barreto de Miranda Sargento

Professora Auxiliar, Universidade de Aveiro (Orientadora)

Professor Doutor Pedro Alexandre de Sousa Goncalves

Professor Adjunto, Universidade de Aveiro (Arguente Principal)

Agradecimentos À Susana pela sua paciência.

Palavras-Chave

Latência, Ponto-a-Ponto, End-to-End, E2E, Round-Trip Time, RTT, Segmentação de Rede, Core, GPRS, EPC, Garantia de Serviço, Testes Activos, 3G, WCDMA, HSPA, 4G, LTE, Monitorização, QoS, Experiência do Utilizador

Resumo

As redes 3GPP de banda larga móvel têm vindo disponibilizar velocidades de transmissão cada vez mais elevadas. Com a introdução das modernas redes de 4G/LTE essas velocidades situam-se agora nos 150 Mbps. Ao mesmo tempo, a saturação no número de subscritores e a facilidade com que estes podem mudar de provedor de serviço, fez com que a retenção dos mesmos fosse algo da maior importância para o sucesso financeiro de um operador móvel.

A atenção dada à experiência de serviço, em todos os seus quadrantes, indo do humano ao técnico, exigiu aos operadores que tenham forma de sentir, quase em tempo real através de sistemas de garantia de serviço, o alinhamento das expectativas dos clientes com a qualidade que a rede disponibiliza. Um parâmetro de rede que está intimamente associado à qualidade da rede e que não é de todo simples nem directo de medir é a latência introduzida por cada um dos elementos ou troços da rede. A importância deste parâmetro de qualidade de rede tem também vindo a ser reconhecida por departamentos de gestão de topo, como já o era pelos departamentos técnicos.

Nesta dissertação explora-se um método que permite ao operador aferir as latências observadas na sua rede, com as suas características únicas de túneis de GTP, utilizando tráfego gerado pelos seus próprios utilizadores, não só de uma forma extremo a extremo (E2E), mas também parcelar. Consegue-se assim obter uma visão geral mas também microscópica da rede, o que possibilita identificar segmentos de rede que estejam a ter impacto na qualidade global da rede, degradando a experiência de utilização. A granularidade permitida possibilita lançar luz sobre segmentos de rede de onde não é de todo habitual conseguir-se sentir o seu comportamento, como é o caso dos segmentos de *core*. A mesma solução irá ainda permitir que os dados recolhidos possam alimentar diversas ferramentas de apoio à decisão, colocando-lhe à disposição uma melhor definição.

De forma a validar a arquitectura proposta nesta dissertação, foram realizados diferentes tipos de testes, em redes comerciais europeias, que possibilitaram validar a relevância dos dados recolhidos, tanto a nível técnico como comercial, e alcançar um melhor entendimento sobre a rede e os seus impactos para a experiência do utilizador final.

Keywords

Latency, Point-to-Point, End-to-End, E2E, Round-Trip Time, RTT, Network Segmentation, Core, GPRS, EPC, Service Assurance, Active Tests, 3G, WCDMA, HSPA, 4G, LTE, Monitoring, QoS, Customer Experience

Abstract

The 3GPP mobile broadband networks have been delivering every time higher throughputs. With the introduction of the modern 4G/LTE networks, those throughputs reached 150 Mbps. At the same time, with the saturation on the number of mobile subscribers and the simplicity on how they change from service provider, made from the retention of the subscribers extremely important for the financial success of a mobile operator.

Due to the attention given to the service experience, in all their quadrants, going from the human to the technical, it demanded from the operators to have a way to feel, almost in real time, by the use of service assurance systems, an alignment of the customer expectations with the quality provided by the network. One network parameter which is deeply related with the quality of service and that is not easily or straightforward to measure is the latency introduced by each of the network segments. The importance of this network quality parameter has been recognized by the top management departments, like it was already by the technical ones.

In this thesis we explore an innovative method which allows the operator to assess the latencies seen in his network, with their unique characteristics of GTP tunnels, using traffic generated by their own users, not only in and E2E approach, but also by each network segment. This way, it is possible to obtain a general view, but also microscopic of the network, which allows to identify network segments which are having impact on the overall network quality, degrading the customer experience. The granularity reached allows to cast light over network segments from which it is not usual to get their behavior, as it is the case from the core segments. The same solution will still permit to feed decision support tools with the data collected, allowing them to have an even higher definition.

In order to validate the architecture proposed in this thesis, we have run different tests, in European commercial networks, which allowed validating the relevance of the data collected, both technical and commercial, and achieving a better understanding about the network and their impacts to the end-user experience.

Índice

Acrónimos

3

3GDT - 3G Direct Tunnel
3GPP - 3rd Generation Partnership Project

A

APN - Access Point Name

C

CAGR - Compound Annual Growth Rate
CAPEX - Capital Expenditure
CDF - Cumulative Distribution Function
CPU - Central Processing Unit

D

DNS - Domain Name System
DSCP - Differentiated Services Code Point
DPI - Deep Packet Inspection

E

E2E - End-to-End
EPC - Evolved Packet Core

G

GGSN - Gateway GPRS Support Node
GPRS - General Packet Radio Service
GSM - Global System for Mobile Communication
GPS - Global Positioning System
GTP - GPRS Tunnelling Protocol

H

HSPA - High Speed Packet Access
HW - Hardware

I

ICMP - Internet Control Message Protocol
IPv4 - Internet Protocol version 4
ISP - Internet Service Provider

L

LTE - Long Term Evolution

M

MBH - Mobile Backhaul
Mbps - Megabits per second
MS - Mobile Station
MT - Mobile Terminal

N

NIC - Network Interface Card

O

OPEX - Operational expenditure

P

PDN - Packet Data Network
PDP - Packet Data Protocol
PGW - PDN Gateway
POP - Point-of-Presence
PS - Packet-Switched
PSTN - Public Switched Telephone Network

Q

QCI - QoS Class Identifier
QoS - Quality of Service

R

RAB - Radio Access Bearer

S

SGSN - Serving GPRS Support Node
SGW - Serving Gateway
SLA - Service Level Agreements
SMS - Short Message Service
SW - Software

T

TCP - Transmission Control Protocol
TE - Terminal Equipment
TFT - Traffic Flow Template

U

UE - User Equipment

W

WCDMA - Wideband Code Division Multiple Access

Y

YoY - Year over Year

Índice de Figuras

Índice de Fórmulas

1. Introdução

1.1 Motivação

A qualidade de rede de banda larga móvel tem vindo a ganhar relevo nos últimos anos, sendo já assumida como um assunto chave na estratégia de evolução apresentada aos investidores [1]. A percepção do utilizador final da qualidade de rede e do serviço de banda larga móvel disponibilizado determina a satisfação do cliente, o que é crucial para o sucesso económico de um operador móvel. Um estudo levado a cabo pelo departamento de *Consumer lab* da Ericsson mostrou que um dos elementos essenciais para a satisfação de um cliente é a velocidade disponibilizada pela rede, ficando este em segundo lugar. À sua frente ficou somente a cobertura geográfica. O valor cobrado pelo serviço e a qualidade do áudio da chamada de voz vem em terceiro e quarto lugar respectivamente [2].

O investimento financeiro que um operador móvel tem que efectuar para adquirir novos clientes tem vindo a crescer nos últimos anos [3] fixando-se entre os €100 e os €300 [4], com impactos directos nos resultados financeiros dos operadores como o EBITDA [1]. Isto faz da retenção dos actuais subscritores, crucial para a rentabilidade da empresa.

Nos últimos anos têm-se observado três tendências a nível mundial no que diz respeito às redes móveis de telecomunicações.

- A explosão do número de subscrições de Banda Larga, que é visível na Figura 1.
- O aparecimento do serviços *Over the Top* (OTT), os quais se servem das redes de banda larga para estabelecer a conectividade ao serviço, podendo desta forma não fornecer qualquer receita ao operador. Isto é o que acontece se o operador não tiver a sua rede preparada para detectar e diferenciar este tipo de serviços dos seus próprios serviços.
- A "Rede" tem cada vez mais vindo a ser utilizada como diferenciadora, principalmente evidenciando a sua qualidade em campanhas de marketing de forma a atrair novos clientes [2].

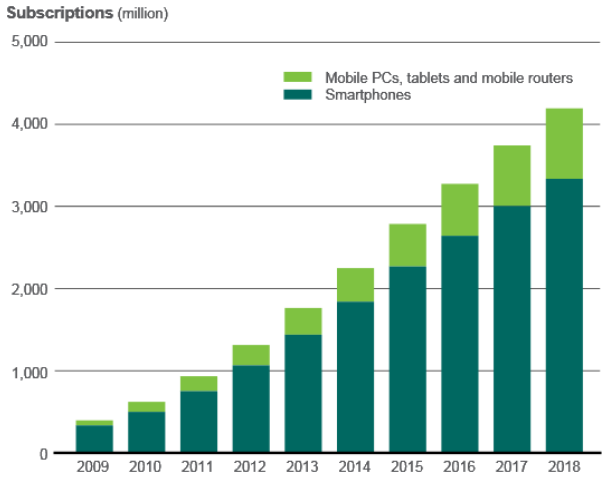

Figura 1 – Evolução no número de subscrições com ligações celulares 2009-2018, em *Smartphones*, *Mobile* PC's, *Tablets* e *Routers* Móveis [2]

A qualidade de serviço de banda larga prestada pelos operadores móveis, medida por entidades independentes, tem também vindo a ganhar relevância nos *media*, tendo os resultados destes *benchmarks* vindo a pesar cada vez mais na altura dos utilizadores se decidirem aquando da aquisição deste tipo de serviço [2]. Os *benchmarks* internos têm também vindo a ser utilizados por operadores móveis transnacionais [1] para colocarem as suas várias participações nacionais em competição de forma a mostrarem à sede qual das participações se distingue, prestando o melhor serviço aos seus clientes, mas ao mesmo tempo a prepará-los para os testes de aferição de qualidade realizados pelos vários reguladores nacionais.

De realçar que ao nível da administração dos operadores existe o foco em melhorar parâmetros de rede que têm influência sobre a qualidade de serviço. Um desses parâmetros é a latência, tendo sido identificada como o quarto parâmetro de rede que necessita ser melhorado [5] com 9% de respostas. A primeira é o *throughput* com 49%, o qual depende largamente do parâmetro referido anteriormente.

Em 2009, em Portugal a Autoridade Nacional de Comunicações (ANACOM) efectuou um estudo [6] que tinha como objectivo aferir a qualidade de serviço em redes de banda larga, o qual incluía pela primeira vez a aferição da qualidade da banda larga em redes móveis. Até esta data, só eram efectuados testes de qualidade na rede fixa. Evidenciava-se neste estudo a mudança de paradigma que se tinha vindo a verificar na definição de qualidade de serviço, desde que se começaram a introduzir novos e serviços cada vez mais complexos, que corriam sobre a banda larga, tanto fixa como móvel. No passado a disponibilidade de acesso e as velocidades de *uplink* e *downlink* alcançadas eram os principais Indicadores Chave de Desempenho (*Key Performance*

Indicators – KPI(s) nos quais o estudo incidia. Actualmente adicionaram-se KPIs que se focam mais na própria experiência do serviço, como a perda de pacotes IP, o tempo que demora a carregar uma página *Web*, mas também o atraso de comunicação, o qual também se pode chamar de latência.

Para ir ao encontro do nível da qualidade de serviço ambicionada pelos clientes, os operadores móveis têm que ter formas de monitorizar em tempo real os parâmetros de rede que têm influência directa na qualidade que a sua rede disponibiliza. Existem soluções no mercado à disposição dos operadores móveis, que dão pelo nome de Garantia de Serviço (do inglês, *Service Assurance*), que monitorizam estes parâmetros de uma forma contínua, numa perspectiva ponto-a-ponto, também referido de E2E, ou então parcelar, sendo esta última mais vista na parte de acesso de rede, entre os NodeB's e os RNC's. As soluções que se focam na percepção final dos utilizadores, que são as já referidas soluções E2E, simulam de uma forma encadeada vários serviços habitualmente utilizados pelos clientes, entre os quais o *browsing* na Internet ou descarregar um ficheiro de um servidor. As soluções que permitem dar uma perspectiva parcelar da rede recorrem sempre à injecção de um fluxo de tráfego com um certo padrão num determinado ponto da rede, e ou recolhem esse mesmo tráfego num outro ponto distante do primeiro para se poder aferir certos parâmetros de rede, ou então o tráfego injectado é reflectido no ponto distante de novo para o ponto de origem, como é o caso do protocolo TWAMP (*Two-Way Active Measurement Protocol*) definido pelo RFC 5357. A grande desvantagem destas soluções é que requer que seja possível encaminhar o tráfego entre o ponto de partida e chegada do tráfego.

As partes da rede cobertas pelas ferramentas de Garantia de Serviço estão indicadas na figura seguinte, onde as linhas contínuas indicam as partes da rede de onde os métodos tradicionais recolhem medidas e que permitem então observar o comportamento agregado da rede, enquanto que as linhas a tracejado indicam os segmentos da rede onde as soluções actuais não conseguem recolherem dados.

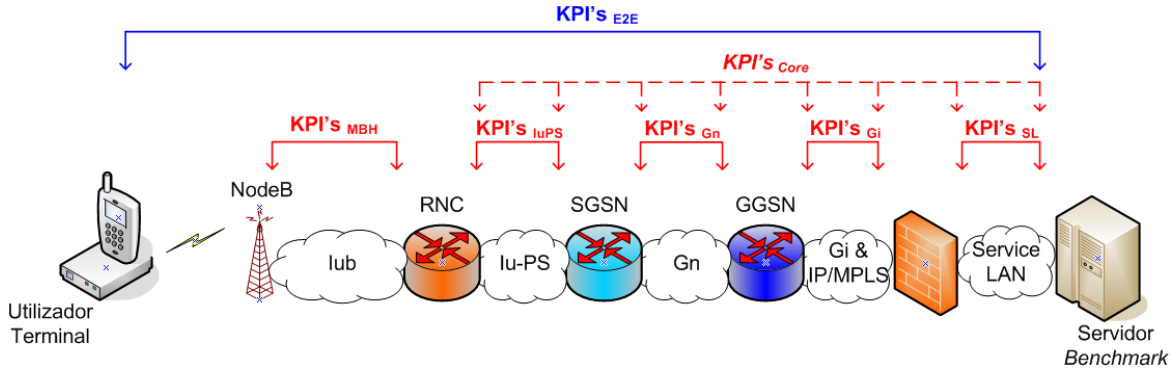

Figura 2 – Alcance dos actuais Métodos de Medida

Estas duas formas de recolha de informação sobre a qualidade de rede não são suficientes a um operador para que consiga identificar de uma forma precisa a parte da rede que necessite, por exemplo, de ser optimizada ou que esteja a ser afectada por um problema num determinado instante.

Estas soluções são claramente mais vocacionadas para redes de banda larga fixa pois, devido a uma singularidade das redes móveis, como o GSM/WCDMA e LTE, nestas existem segmentos de rede onde são utilizados túneis que interligam elementos de rede como por exemplo, o RNC ao SGSN, o SGSN ao GGSN ou o eNodeB ao SGW, os quais não atravessam os referidos elementos de rede, logo não sendo possível encaminhar tráfego entre estas redes, a não ser utilizando o túnel entre o UE e a *Gateway*, como indicado na Figura 13. Esta particularidade faz com que as referidas soluções de monitorização não consigam cobrir na plenitude a aferição da qualidade de rede, pois colocam de fora das medições, os elementos de rede onde haja pelo menos uma terminação de um dos referidos túneis.

Esta dissertação propõe-se expor um método inovador, simples, financeiramente interessante e preciso que permite a operadores móveis terem a visibilidade necessária sobre a sua rede da banda larga móvel 2G/3G ou 4G, não só de uma perspectiva E2E, que permite determinar a percepção do cliente final, mas também decompondo por cada segmento da sua rede os KPIs que a caracterizam, tanto em *downlink* como em *uplink*, e que definem e determinam a qualidade de uma rede.

Com um sistema destes integrado na rede e interligado aos seus sistemas de qualidade de rede e de alarmística, um operador móvel poderá detectar problemas na rede ao nível do serviço prestado e actuar de uma forma proactiva, evitando que a degradação do serviço seja percepcionada pelo seu cliente final. O referido método recorre a testes activos, iniciados por

utilizadores de teste, sendo este tráfego capturado em pontos específicos da rede, dependendo unicamente da granularidade desejada pelo operador.

Denotar que, por comodidade e simplicidade, os já referidos *benchmarks* recorrem habitualmente a pedidos de ICMP, a que vulgarmente de chamam de PING's, para determinar o valor de *Round Trip Time* (RTT) da rede. O PING como ferramenta é mais adequado para fazer *troubleshooting* de conectividade. Permite determinar um valor aproximado de RTT entre dois pontos, mas não é efectivamente a ferramenta ideal para determinar com exactidão a verdadeira latência que existe entre dois pontos. Nesta dissertação iremos também explicar porque é que o valor medido por PINGs não é relevante para se tentar aferir a qualidade da rede móvel, pois a forma como é medida, devido às característica das redes móveis, não afere com exactidão um valor que poderia indiciar a experiência de serviço disponibilizada pela rede.

Tendo a latência de rede uma importância crucial na velocidade máxima de *download* e *upload* que se pode alcançar, também em aplicações de Voz e em Jogos *Online*, será neste KPI que esta dissertação se irá concentrar.

O interesse e a urgência que um operador móvel tem em começar a se aperceber do comportamento da sua rede em termos de latência, provém do facto de que, com o salto para redes 4G/LTE/EPC e com a introdução na rede de *core*, tanto de redes SDN (*Software-Defined Networking*), que permitem a abstracção do *Software* da plataforma onde ele corre, como de redes SON (*Self-Organizing Network*) [7], que proporcionam ao operador obter ganhos operacionais devido à simplicidade com que as redes são autogeridas, é de extrema importância ter esta visibilidade de uma forma permanente, pois se os elementos ou segmentos de rede que não tenham um comportamento previsível, dentro de limites definidos pelo operador, podem por em causa a experiência de utilização, logo passando uma imagem negativa do serviço prestado pela rede.

Com a solução que se desenvolveu, pretendemos passar de uma abordagem próxima de uma caixa-negra, como a indicada na Figura 3, que é a forma que actualmente se encontra nas soluções de garantia de serviço disponíveis no mercado, para uma abordagem mais ampla, com maior granularidade, que é sucintamente representada pela Figura 4.

Iremos argumentar que enriquecendo as soluções de garantia de serviço com o método descrito, se consegue entregar mais valor ao operador móvel do que ele consegue retirar com os métodos actuais.

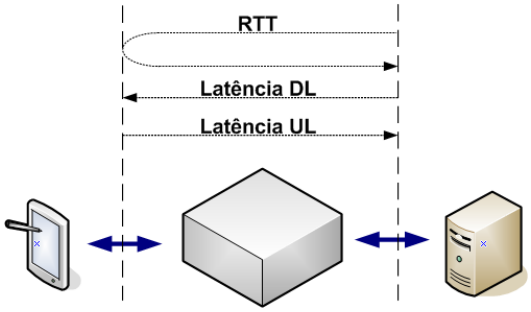

Figura 3 – Aferição da qualidade de Rede numa abordagem E2E tipo Caixa-Negra

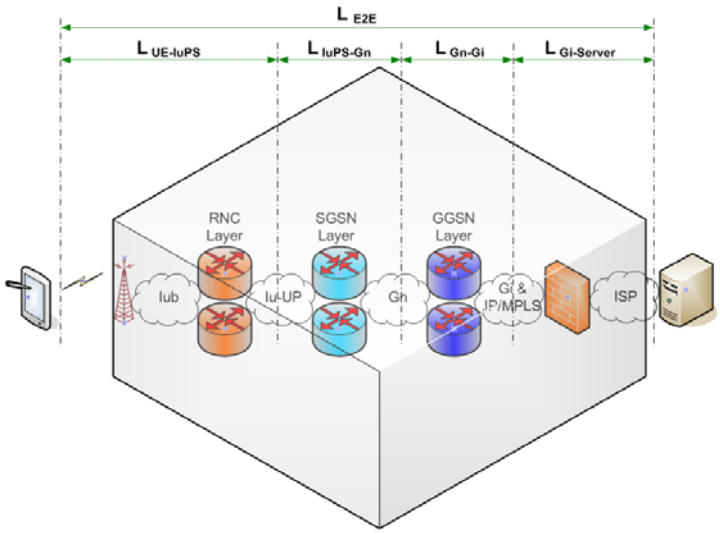

Figura 4 – Aferição da qualidade de Rede numa abordagem mais Granular

1.2 Objectivos

O objectivo principal desta dissertação é descrever um sistema de garantia de serviço que permite a operadores móveis aferirem a latência que cada segmento da sua rede introduz numa sessão de dados. Irão ser utilizados testes activos com tráfego de cliente reais. Iremos explicar a solução que se desenvolveu, tanto ao nível de *hardware* como *software*. Existem no entanto alguns constrangimentos, pois é uma solução desenvolvida com o apoio de um fornecedor de redes móveis, o que não nos permite expor na sua plenitude a mesma. No entanto iremos evidenciar grande parte das suas características.

Um protótipo deste sistema foi utilizado em redes comerciais de dois Operadores europeus, um *Tier-1* e um *Tier-2*, nos quais foram recolhidos dados das suas redes 3G de banda larga, e analisados os seus valores de latência. Este parâmetro caracteriza a rede mas, no entanto, não existe uma forma directa de o recolher através de estatísticas dos elementos que constituem a

rede, e no entanto é um dos que mais intimamente está relacionado com a experiência de serviço do utilizador.

Iremos mostrar os valores de latência, de uma forma E2E, entre o servidor de *benchmark* e o UE (do Inglês, *User Equipment*), mas também a contribuição que cada segmento de rede tem, como por exemplo entre o Iu-UP e a Gn, para esse valor final agregado. De referir que esta visão parcelar e integrada da latência é o que diferencia a nossa solução das restantes soluções existentes actualmente no mercado. Estes valores podem então ser utilizados como valores de referência para outras redes.

Por fim iremos também apresentar casos práticos que mostram que a solução desenvolvida consegue identificar eventos no núcleo da rede que à data desta dissertação, com as soluções existentes, passariam completamente despercebidos ao operador, e que no entanto estes eventos podem ter um impacto altamente negativo na experiência de utilizador.

1.3 Contributos deste Trabalho

Com a solução proposta e com a realização dos objectivos acima enunciados pretende-se que este trabalho forneça as seguintes contribuições.

- Desenvolvimento de uma solução que permita a fornecedores de soluções de *Service Assurance* enriquecerem as suas propostas para redes móveis, ultrapassando os constrangimentos impostos pela arquitectura utilizada em redes móveis de banda larga, e conseguindo extrair desta o indicador da latência, que caracteriza de sobre maneira a experiência de utilização de um cliente.
- Validar em redes móveis que se encontram ao serviço, que os dados recolhidos identificam o desempenho sentido pelo cliente e que consegue caracterizar segmentos de rede que, à data de hoje um operador móvel, não tem forma de validar o seu comportamento.

1.4 Organização do Projecto

Até ao momento, este documento disponibilizou uma breve introdução e uma descrição no que se espera que seja a contribuição desta dissertação. Nos próximos capítulos iremo-nos focar no seguinte.

- **Capítulo 2**

Estado do mercado da banda larga móvel, da tecnologia de rede envolvida e de alguns conceitos genéricos que ajudarão a compreender melhor o porquê da urgência e necessidade de uma solução deste género, com especial atenção na experiência de serviço.

- **Capítulo 3**

Iremos fazer uma breve introdução à arquitectura de redes móveis de banda larga, como definida pelos 3GPP. Descreveremos os elementos de rede mais relevantes, as redes existentes e por fim falaremos de como as sessões de dados são activadas e como o seu fluxo de dados pode ser controlado.

- **Capítulo 4**

Será utilizado para identificar os principais parâmetros que caracterizam as redes e como influenciam a experiência do utilizador.

Iremos também descrever o que é a latência ou atraso de rede. Iremos descrever os dois tipos que existem nas redes móveis e qual o seu impacto para o utilizador final. Argumentaremos sobre as desvantagens da utilização de pedidos de ICMP, para aferir a latência numa rede móvel. Por fim, falaremos sobre as diversas possibilidades de recolha de estatísticas da rede.

- **Capítulo 5**

Iremos iniciar a descrição da solução implementada, a sua arquitectura, como é feita a captura do tráfego, o armazenamento dos dados recolhidos e como é efectuada a análise da informação. Iremos também posicionar a solução no universo das soluções de garantia de serviço.

- **Capítulo 6**

Apresentação dos cenários de testes efectuados e qual o racional por detrás de cada um.

- **Capítulo 7**

Neste capítulo iremos apresentar os resultados obtidos nos diversos testes e discutir sobre os mesmos.

- **Capítulo 8**

Por fim, apresentaremos as conclusões e o que pretendemos desenvolver num futuro muito próximo.

2. Sistemas de Garantia de Serviço

2.1 Introdução

Desde que começaram a ser introduzidos novos serviços pelos operadores móveis, como aconteceu em meados da década 1990 com a introdução dos serviços pré-pagos e de SMS (em Inglês, *Short Message Service*), que os sistemas de apoio ao negócio começaram a ter uma relevância considerável nas empresas de telecomunicações. O aparecimento dos serviços de dados como o GPRS que permitiu por sua vez oferecer serviços de valor acrescentado aos subscritores das redes móveis, implicou uma mudança maior nos referidos sistemas, pois as unidades de negócio dos operadores móveis tiveram necessidade de obter uma visibilidade mais profunda sobre a percepção que os seus clientes tinham dos vários serviços que estavam ao seu dispor.

Actualmente, nos países com uma história mais longa de serviços móveis de telecomunicações, como é o exemplo de Portugal, verifica-se uma saturação no mercado das subscrições móveis [8], direccionadas para as pessoas. No entanto, no universo das subscrições móveis para comunicações máquina-para-máquina (M2M, do Inglês, *Machine-to-Machine*) prevê-se ainda uma elevada margem de desenvolvimento nos próximos anos, em que existe a visão dos 50 mil milhões de dispositivos ligados até 2020 [9]. Esta última tecnologia tem impacto em diversos segmentos como por exemplo, a telemática, a localização de frotas ou *Smart Grids* e tem a vantagem de poder ser utilizada por mais do que uma indústria como é o exemplo da indústria automóvel, empresas financeiras, de energia ou saúde.

A referida saturação das subscrições direccionadas para as pessoas faz com que os operadores móveis tenham que ser mais criativos de modo a preservarem a sua cota de mercado. Para conseguirem endereçar este desafio têm que estar especialmente atentos às situações de risco que advêm do desalinhamento entre a qualidade de serviço que os clientes esperam ou necessitam e a qualidade que a sua rede disponibiliza. É nesta área de percepção e antecipação de risco de perder clientes (em Inglês, *churn*) que sistemas específicos de apoio ao negócio, como os sistemas de garantia de serviço, são utilizados.

Na literatura [10] definem-se habitualmente as seguintes 5 grandes áreas, que se encontram organizadas de forma a permitir facilmente enquadrá-las nas várias áreas de acção dos operadores.

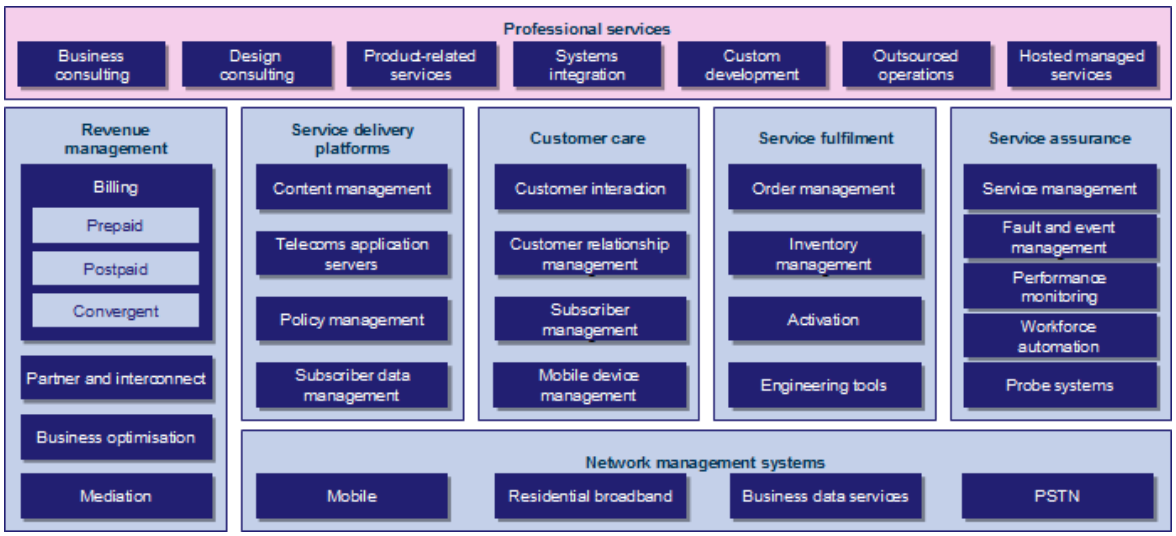

Figura 5 – Segmentos dos diversos mercados de software para Operadores de Telecomunicações [11]

Estas vão desde a gestão de receita às plataformas de prestação de serviços, passando pelas plataformas de atendimento a clientes e de desempenho de serviço, por fim encontram-se as plataformas de garantia de serviço. Cada uma delas subdivide-se ainda mais, descriminando de uma forma mais detalhada as reais necessidades de um operador moderno de redes móveis. De seguida iremos descrevê-las de uma forma sucinta.

- **Sistemas de Gestão de Receita**

Neste segmento definem-se todos os sistemas de software que gerem tanto os pagamentos que o operador tem que efectuar a parceiros como também às recolhas de receitas que lhes são devidas.

Aqui encontramos os sistemas como os de taxação de pré-pago e pós-pago, os de pagamentos devidos a parceiros de *roaming* ou a terminações de chamadas de voz, os de cobrança a parceiros de entrega de conteúdos. É também neste segmento que se encontram definidos os sistemas de garantia de receita e de gestão de custos.

- **Plataformas de Entrega de Serviço**

São componentes de *software* que permitem incrementar a velocidade com que se desenvolve, entrega e se gerem, tantos os de serviço actuais como os que se encontram em desenvolvimento.

Como exemplos temos serviços de localização, serviços de toques de telemóvel, serviços de informação sobre subscritores e por fim serviços de gestão de políticas, direccionadas aos tipos de serviço que o cliente utiliza.

- **Sistemas de Atenção ao Cliente**

Este tipo de sistema permite ao operador tratar e controlar a forma como os seus clientes interagem com ele, tanto através de atendimentos personalizados ou automatizados. Permite também oferecer serviços adaptados ao perfil do cliente e registar os seus pedidos ou reclamações.

Têm o intuito de aumentar a satisfação do cliente em situações que possam ser consideradas de algum descontentamento sobre o serviço, diminuir possíveis desistências de subscrições, reduzir o custo com o centro de atendimento e potenciar o aumento de receita com ofertas mais adequadas ao perfil do cliente.

Aqui encontram-se sistemas como o de Gestão de Relacionamento com o Cliente (em Inglês, *Customer Relationship Management* ou CRM), os sistemas de interacção com o cliente como o centro de atendimento ao cliente, portais *self-service* ou atendimentos de voz automáticos (do Inglês, *Interactive voice response* ou IVR).

- **Sistemas de Entrega de Serviço**

São plataformas que permitem ao operador móvel gerir o processo de ordens de serviço, de inventário e activação de serviço. Inclui também todas as ferramentas utilizadas pelos departamentos de engenharia para operar de uma forma mais eficiente.

- **Sistemas de Garantia de Serviço**

A solução descrita nesta dissertação posiciona-se nesta área, enquadrando-se na área mais à direita da Figura 5, e irá ser descrita no capítulo seguinte de uma forma mais exaustiva.

2.2 Definição

Os sistemas de garantia de serviço existem para apoiarem as unidades de negócio do operador móvel extrair dados dos vários elementos de rede, monitorizando elementos desde o acesso móvel, como as BTS, NodeB's e eNodeB's, passando pelas camadas de transporte e pelo *Core* IP e *Packet Core*. No entanto, sistemas mais modernos conseguem observar o próprio serviço que o cliente está a utilizar, podendo alcançar a camada 7, ou seja a camada aplicacional do modelo OSI.

Todos os fluxos de dados recolhidos têm que ser processados em tempo real, de forma a fornecer ao operador informação relevante que possa então ser utilizada para aferir, tanto a qualidade do serviço prestado como a experiência do próprio cliente. De forma a enriquecer a análise, quantos mais fluxos de dados o operador móvel tiver à sua disposição, mais pormenorizada será essa análise e maior probabilidade de sucesso terá a decisão efectuada tendo essa informação como base.

Os sistemas de garantia de serviço subdivide-se nas seguintes sub-áreas.

2.2.1 Gestão de Serviço

Caracteriza os sistemas de software utilizados pelos operadores móveis para acederem a cada um dos seus serviços individualmente, e que lhes permite retirar relatórios granulares, tanto por cliente como por serviço, como também lhes permite aferir a qualidade do serviço prestado.

2.2.2 Falha e Gestão de Serviço

São sistemas que se interligam aos elementos de rede e aos elementos de gestão fornecidos pelos fabricantes. Têm como principal função a de processar e filtrar os eventos da rede de forma a isolar e tentar encontrar a causa de um problema específico.

Têm também a função de alertar a parte da organização responsável pela supervisão da rede sobre situações que não sejam consideradas normais em qualquer parte da rede.

2.2.3 Gestão de desempenho

Recolhem os dados dos vários elementos de rede de uma forma contínua, de forma a permitir ao operador gerir o desempenho actual desses elementos, como também o de permitir verificar

esse mesmo desempenho ao longo do tempo, permitindo desta forma aferir as tendências dos vários indicadores.

2.2.4 Automação da Força de Trabalho

Permite ao operador gerir incidentes resultantes da disrupção do serviço e do envio de recursos para o terreno de uma forma eficaz através da abertura de ordens de serviço, optimizando o conjunto de competências que ajudarão de uma forma mais eficaz na resolução do problema.

2.2.5 Sondas

Estes sistemas são uma combinação de Hardware e Software, em que os elementos de Hardware são colocados em pontos estratégicos da rede, os quais monitorizam de uma forma passiva tanto a sinalização como os dados enviados pelos clientes. Podem também ser distribuídos pela rede, de novo em pontos específicos para testarem de uma forma automatizada diversos acessos, como o 3G ou 4G, ou tecnologias como é o exemplo das interfaces de *ethernet*.

A solução que se descreve nesta dissertação, inserida na área de garantia de serviço, cobre dois dos sub-domínios indicados, o de gestão de desempenho (*Performance Management)* e o de sistema de sondas (*Probing Systems*).

2.3 Soluções do Mercado

No mercado já se encontram várias soluções que cobrem os vários segmentos enunciados no capítulo anterior. De seguida iremos enunciar as soluções de sondas mais em voga actualmente em uso no mercado, e evidenciar o que de mais relevante elas disponibilizam ao operador. Iremos ao mesmo tempo referir algumas soluções de gestão de desempenho dos mesmos fornecedores de sondas que de certa maneira completam essas mesmas soluções.

No capítulo seguinte iremos apresentar o posicionamento da solução descrita nesta dissertação e os pontos que consideramos diferenciadores relativamente ao que o mercado propõe e onde o seu valor pode complementar as soluções já existentes.

2.3.1 Sistemas de Sondas

Existem no mercado 7 grandes fornecedores de sistemas de sondas [11], que no total perfazem mais de 70% de cota de mercado. De seguida iremos agregar a informação mais relevante sobre cada um desses sistemas, subdividindo-os por segmentos de produto, de forma a conseguirmos identificar as principais ideias que regem actualmente as soluções mais largamente aceites pelos operadores.

Incluímos também nesta tabela os sistemas de gestão de desempenho, pois são um complemento essencial aos sistemas de sondas, pois permitem a visibilidade dos dados recolhidos pelos últimos.

Tabela 1 – Ofertas no Mercado – Tektronix

Designação da Solução	Segmento	Descrição
GeoProbe (G10)	Sonda	Sistema de sondas passivas distribuídas pela rede que permite a recolha de dados em tempo real. Fornece recursos de garantia de rede para operações de rede incluído, monitorização proactiva, diagnostico, despiste de problemas de serviço e análise protocolar, mais orientada para os fluxos de sinalização. É uma solução independente dos fornecedores dos elementos de rede.
Iris Suit	Gestão de Desempenho	Solução que efectua despiste de problemas de uma forma automatizada. Permite extrair informação adicional dos serviços utilizados pelos clientes. Permites também alertar o operador quando um determinado KPI ultrapassa um determinado valor pré configurado. Desta forma o operador pode pró-activamente tentar resolver o problema antes de o cliente dar conta da degradação excessiva da qualidade de serviço.

Tabela 2 – Ofertas no Mercado – JDS Uniphase

Designação da Solução	Segmento	Descrição
NetComplete Ethernet	Gestão de Desempenho	Conjunto de produtos de garantia de serviço para despiste de problemas, gestão de performance, acordo de nível de serviço (em Inglês, *Service Level Agreement* ou SLA) e de relatório. É uma solução independente dos fornecedores dos elementos de rede.
RCATS Remote Test Probes	Sonda	Sondas de teste colocadas em locais relevantes que simulam os serviços utilizados pelos clientes do operador. Permite monitorizar activamente e medir a qualidade de serviço de uma perspectiva unicamente E2E.
QT-6x0 Ethernet Probes	Sonda	Sondas de teste colocadas em interfaces *Ethernet* da rede do operador, que permitem efectuar testes activos, nos quais a sonda injecta tráfego na rede tentando simular o tráfego de um cliente tipo, ou de uma forma passiva, na qual o operador configura filtros para capturar mensagens consideradas relevantes.

Tabela 3 – Ofertas no Mercado – NetScout

Designação da Solução	Segmento	Descrição
nGenius Probes	Sonda	Permite visibilidade em tempo real das camadas 2 à 7 do modelo OSI do tráfego capturado, permitindo assim caracterizar os serviços que flúem na rede e permitindo ao mesmo tempo alertar o operador se um determinado KPI ultrapassar um valor pré estabelecido.
nGenius Performance Manager	Gestão de Desempenho	Permitem monitorizar todo o tráfego capturado na rede pelo sistema nGenius, disponibilizando também o histórico do volume e a resposta de serviço tanto no plano de sinalização e de tráfego do cliente. Permite visibilidade por subscritor.

Tabela 4 – Ofertas no Mercado – Empirix

Designação da Solução	Segmento	Descrição
X-EMS	Sonda	Permite compreender a experiência do cliente quando acede a serviços de dados móveis através de monitorização E2E da rede.
xCentrix	Sonda	Solução que permite uma visibilidade sobre como os utilizadores estão a utilizar os recursos de rede, observar as tendências e ajudando o operador a tomar a melhor decisão de investimento. Permite também pró-activamente diagnosticar os serviços analisando os eventos registados na rede e quais as evoluções de utilização dos mesmos.

Tabela 5 – Ofertas no Mercado – Astelia

Designação da Solução	Segmento	Descrição
Ocean	Sonda	Sistema que monitoriza a qualidade de redes móveis, de uma forma E2E, de uma perspectiva do utilizador.
Trending and Aggregation	Gestão de Desempenho	Fornece KPI's da rede e análise de tendências.

Tabela 6 – Ofertas no Mercado – Accedian

Designação da Solução	Segmento	Descrição
V-NID Suite	Sonda	Permite recolher medições de latências, num sentido, nos dois sentidos, perda de pacotes, pacotes desordenados, *Jitter*, etc. Para obter estes valores, a solução injecta tráfego gerado pelas sondas entre os dois pontos de rede que se pretende medir. Tem a vantagem de utilizar a tecnologia estandardizada *Two-Way Active Measurement Protocol* (TWAMP) definida em [12].
	Gestão de Desempenho	Fornece KPI's da rede e análise de tendências. Permite alertar as operações quando um determinado limite de um KPI é alcançado.

Tabela 7 – Ofertas no Mercado – Cisco

Designação da Solução	Segmento	Descrição
IP SLA	Sonda	Parte do Software IOS que corre nos *routers* do mesmo fornecedor. Utiliza uma monitorização activa de tráfego, gerando tráfego entre os dois pontos da rede que se pretende medir. Permite analisar níveis de serviço da camada IP.

Filtrando o que foi recolhido nos diversos *Data Sheets* dos fornecedores, podemos verificar que existem os seguintes 7 grandes vectores:

- Soluções baseadas em Interfaces *Ethernet*;
- Independentes dos Fornecedores dos Elementos de Rede. Isto é, Soluções baseadas em HW/SW externo aos elementos de rede ou então SW que se encontra incorporado no CPU do elemento de rede. A única solução que se encontra neste último campo, incorporada no SW do elemento de rede, é a solução IP SLA da Cisco;
- Monitorização da Qualidade de Rede de uma perspectiva E2E ou então entre pontos da rede entre os quais existe conectividade, isto é, entre os quais se podem efectuar pedidos de ICMP (PING's) nos quais existe resposta;
- Visibilidade por Utilizador;
- Despoletar Alertas se um determinado KPI atingir um valor predefinido;
- Proactividade na resolução de problemas;
- Monitorização em Tempo Real.

Investigando um pouco mais o segundo vector, sobre soluções baseadas em HW/SW, nas quais se recorre a equipamento dedicado, ou baseadas somente em SW, que utilizam módulos de SW instalados no CPU de um *router* de IP, por exemplo, fez-se a seguinte análise sobre as duas abordagens. Nesta análise tivemos em conta os aspectos técnicos mas entramos em conta com os possíveis impactos financeiros das duas possibilidades, no relatório financeiro de um operador móvel de telecomunicações.

Começando pelas soluções baseadas em HW/SW.

- **Soluções baseadas em HW/SW – Prós**
    - Medições de alta Precisão. Podem recorrer a equipamento de sincronismo externo como GPS ou IEEE1588 ou então mecanismos proprietários.
    - Cadência entre pacotes elevada, que podem chegar a valores tão baixo como 100 µs.
    - Possibilidade de gerar tráfego diverso (TCP, RTP, ICMP, etc) com marcações de DSCP distintas.
    - Possibilidade de medir pacotes duplicados ou desordenados.
    - Integração com soluções de *Performance Measurement ou Fault Management*.
    - Permitem definir limiares a partir dos quais alertas são gerados e que permitem acções céleres por parte do operador para mitigar ou resolver a situação.

- Custo de operação relativamente baixo, depois de integrado na rede – OPEX.

- **Soluções baseadas em HW/SW – Contras**
  - Custo de aquisição, ou seja CAPEX.
  - Necessidade de integração com a infra-estrutura de Layer 2 (switches) do operador e consequente ocupação de um determinado número de portas neste equipamento.
  - Necessidade de injecção de tráfego não autóctone na rede do operador, isto é, que não é gerado por um subscritor seu.
  - Custos de Suporte associados ao equipamento, entrando então nas despesas operacionais do operador, isto é OPEX.
  - Não consegue aferir o impacto de elementos de rede como SGSN, GGSN, SGW ou PGW, devido à existência dos túneis de GTP.

Observando agora as soluções baseadas em SW.

- **Soluções baseada somente em SW – Prós**
  - Não existe a necessidade de adquirir equipamento específico para este propósito – CAPEX.
  - Não tem custos operacionais extras para o Operador de pois de estar em serviço, se não necessitar configurações novas – OPEX.
  - Integração com soluções de *Performance Measurement ou Fault Management*.
  - Permitem definir limiares a partir dos quais alertas são gerados e que permitem acções céleres por parte do operador para mitigar ou resolver a situação.

- **Soluções baseada somente em SW – Contra**
  - Utiliza capacidade de CPU ao *router* onde está a correr. Depende do número e do tipo de operações configuradas, podendo chegar facilmente a impactos negativos próximos dos 15%.
  - Impactos negativos no CAPEX. Pois pode levar à necessidade de aquisição de novos *router's* dependendo da ambição por parte do cliente no número e tipo de medições a serem efectuadas.
  - Cadências de pacotes a rondar os ms.
  - Somente permite medir *Round Trip Times*.

- Permitem utilizar fluxo de TCP ou UDP.
- Não permitem medir pacotes duplicados ou desordenados.
- Correm sobre Infra-estrutura crítica a qual transporta tráfego que é o *Core Business* do operador. Qualquer impacto negativo neste equipamento pode afectar o serviço prestado ao cliente final.
- Não permitem Visibilidades do Utilizador Final.
- Não consegue aferir o impacto de elementos de rede como SGSN, GGSN, SGW ou PGW, devido à existência dos túneis de GTP.

Pesando ambos os lados da balança podemos pensar que possivelmente as soluções por SW podem ter alguns riscos associados, mas que aferidos correctamente podem ser mitigados. As soluções por HW parecem ser as ideais, no entanto têm um custo associado que, tanto a nível de CAPEX como OPEX, poderá ser difícil defender e suportar perante o CTO do operador.

Em [11] são dadas algumas recomendações aos operadores móveis, entre as quais se encontram, a extensão da cobertura das sondas de IP, para mais perto das extremidades da rede, para mais rapidamente detectar e diagnosticar problemas. É também evidenciada a necessidade de convergência entre os sistemas se voz e de dados, de forma a se implementarem sistemas capazes de cruzar os domínios tecnológicos e fornecer relatórios que permitam uma melhor visibilidade da experiência do utilizador. Sugere-se ainda que se preparem os sistemas de forma a monitorizar o desempenho e as falhas para quando os serviços de *Cloud* se generalizem e que as soluções devam ser abrangentes, de forma a permitirem cobrir um largo espectro de serviços. Por fim recomenda-se que deverá haver um investimento em sistemas de monitorização que permitam uma visão mais detalhada da sessão do cliente.

Iremos demonstrar no capítulo 5 que a forma como a solução descrita nesta dissertação foi desenvolvida, cobre grande parte dos vectores indicados anteriormente e que as complementa com três ideias chave:

- Tornar independente de se ter ou não conectividade entre as várias redes da rede de banda larga definida pelos 3GPP.
- Permitir enriquecer globalmente as soluções de garantia de serviço com propostas de valor mais alinhadas com as necessidades actuais dos operadores moveis de banda larga, que desejam retirar o máximo de visibilidade de como o utilizador percepciona a sua rede, antecipando com a monitorização adequada quais quer queixas que possam vir a resultar de impactos negativos da performance de elementos ou segmentos da sua rede.

- Poder alimentar a solução de *Customer Experience Manager* (CEM) existente no operador, com os dados recolhidos pela solução de modo a enriquecer a qualidade da análise da referida solução.

2.4 Valor Acrescentado pela Solução Proposta

Na figura seguinte indica-se de uma forma mais pormenorizada, como os vários sub-domínios pertencentes à área de garantia de serviço interagem entre elas e quais as peças que fazem parte de cada um desses sub-domínios. A vermelho indicam-se as áreas nas quais a solução proposta nesta dissertação consegue cobrir.

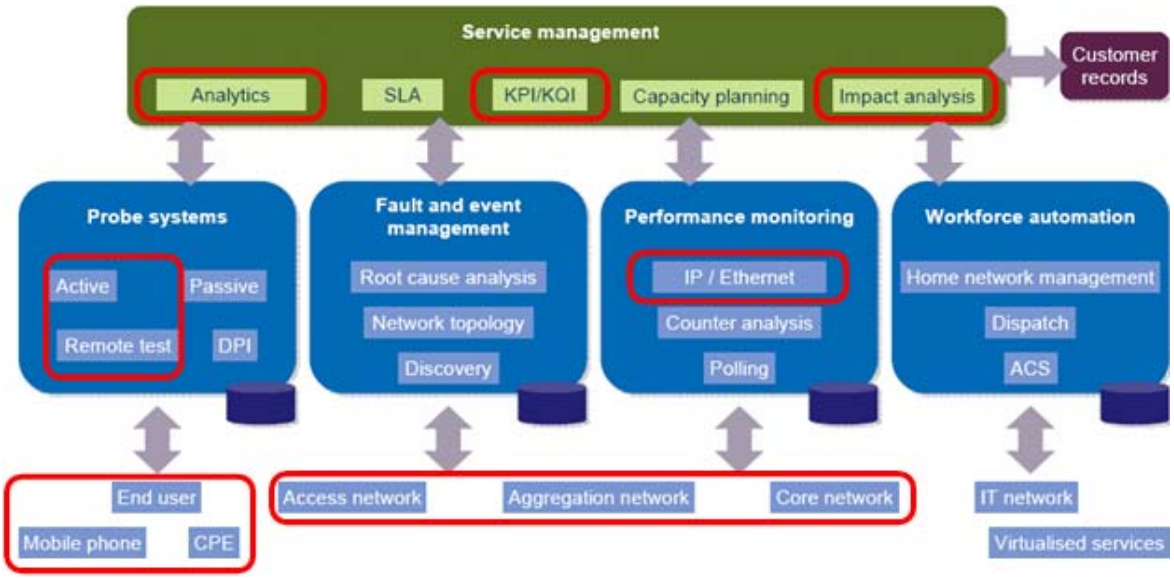

Figura 6 – Garantia de Serviço – Funcionalidades de Software por Sub-Domínio [11]

As propostas de valor da solução são:

- Evolução das *probes* passivas existentes na rede com a funcionalidade de marcação do instante de tempo da captura tráfego:
  - As *probes* passivas adaptadas de forma a efectuarem a marcação do instante de tempo no qual um determinado pacote foi capturado, valor este inserido no pacote de *Ethernet* na parte do FCS, irá permitir que se façam medições extremamente precisas nos cálculos de latência, entre o mesmo pacote, recolhido nas diversas *probes* espalhadas pelos diversas interfaces redes, garantindo que o conjunto de *probes* se encontram sincronizadas, por GPS ou IEEE 1588 (PTPv2).

- Versatilidade
  - Poder-se utilizar o tráfego gerado por sondas remotas espalhadas pela rede gerando tráfego com o intuito de simular um utilizador real.
  - Monitorizar subscritores individuais como por exemplo CxO's do operador ou de parceiros relevantes, podendo desta forma a experiência de serviço que a referida pessoa está a ter, ou então monitorizar grupos tipificados de subscritores, podendo-se assim poupar no investimento em sondas remotas.

3. Redes Celulares de Banda Larga

3.1 O Mercado Actual do Operadores Móveis

A indústria das comunicações móveis está numa fase de mudança de paradigma. Até à segunda metade da primeira década do século XXI as receitas dos operadores móveis estavam intimamente ligadas à utilização dos recursos existentes na sua rede. Por exemplo, quando se efectuava uma chamada, a duração da ocupação dos recursos de rede eram cobradas ao utilizador. Com o aparecimento e posterior massificação da banda larga móvel esta linearidade entre utilização dos recursos e a receita deixou de ser válida. Isto é visível no seguinte gráfico.

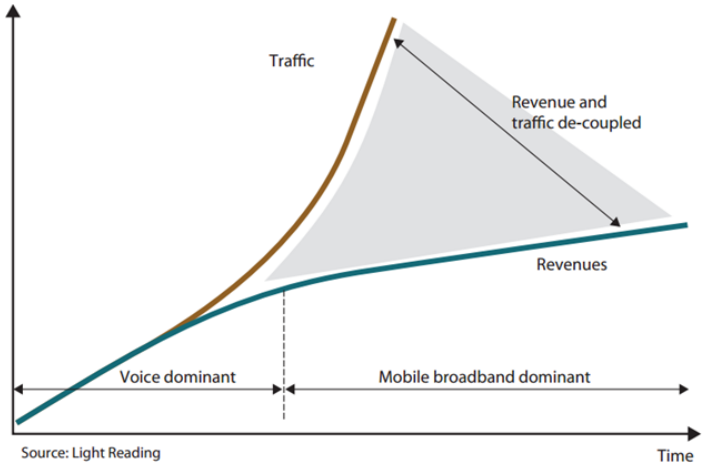

Figura 7 – Divergência Tráfego Processado vs. Recitas

As receitas dos operadores, obtidas com o serviço de voz ("*Voice MTR*"), têm diminuído nos últimos anos, devido principalmente à intervenção dos reguladores [13]. No sentido inverso, as receita com os Dados ("*Data*"), principalmente da Internet Móvel, tem crescido a um ritmo bastante interessante, 20%-47% YoY [13], mas que ainda não tem compensado os operadores pelo decréscimo das receitas na voz. Isto é observável no gráfico seguinte, que apresenta a evolução das receitas anuais, disponibilizado no relatório de resultados do Grupo Vodafone.

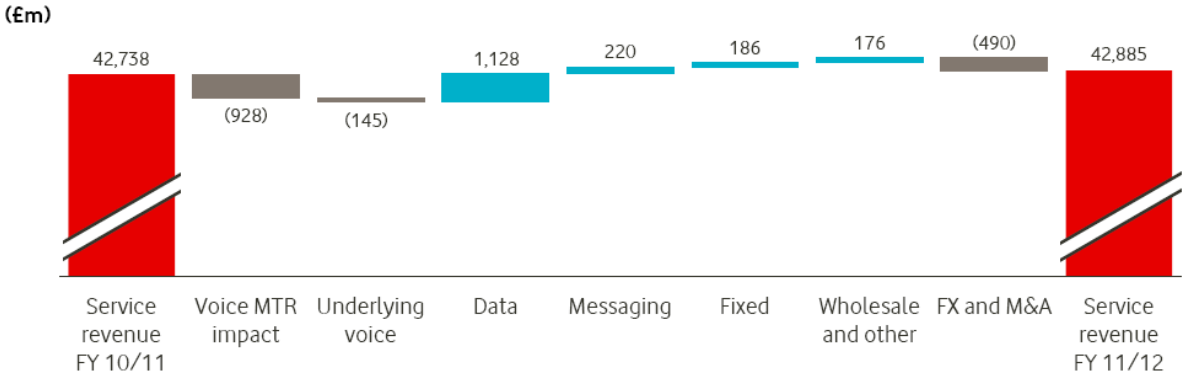

Figura 8 – Exemplo de Evolução das Receitas FY10/11 para FY11/12 [6]

Até 29 de Junho de 2007, um telemóvel era essencialmente um dispositivo para efectuar ou receber chamadas de voz ou para enviar e receber SMS's, o qual também era utilizado esporadicamente para aceder a serviços de dados, como por exemplo WAP ou MMS. A partir da referida data, na qual foi anunciado ao mercado um dispositivo com um sistema operativo para dispositivos móveis que fez mudar hábitos e que conseguiu desviar o epicentro da essência tradicional do telemóvel, que era o de efectuar ou receber chamadas de voz. Esse dispositivo possibilitou a revolução de várias indústrias, desde a de *hardware* à de *software*, catapultando a criação de aplicações que inovaram a forma como as pessoas começaram a comunicar, a socializar e até mesmo trabalhar. Esse dispositivo móvel revolucionário recebeu o nome iPhone e era um produto da Apple. Um ano mais tarde, a 22 de Outubro de 2008 foi lançado um outro sistema operativo, o Android, que permitiu uma concorrência mais plena neste mercado.

Estes dispositivos, a que o marketing empresarial aproveitou um termo de 1997, para lhes dar um nome mais apelativo e diferenciador, *Smartphone*, prevê-se que venham a ter uma evolução galopante a nível mundial nos próximos anos, chegando aos 16%, CAGR, no período 2010 a 2017 [14].

Este crescimento potencia um novo fôlego aos operadores móveis, pois permite-lhes renovar os dispositivos móveis que utilizam a sua rede, colocando alguma fidelização aos utilizadores, e porventura com uma cadência na substituição de terminais mais rápida do que na altura. O racional do operador é que o utilizador deste tipo de terminais tenha outras necessidades, disponibilizadas pelas funcionalidades existentes no terminal, em que ele veja valor na aquisição de outro tipo de pacotes, um pouco mais dispendiosos mas que se tornem essenciais para o uso despreocupado. Este comportamento tem como consequência, um maior ARPU, logo uma maior rentabilidade. No entanto, uma nova realidade veio associada aos *Smarthpones*, e que ameaçou a rentabilidade da indústria das redes móveis, os Serviços *Over the Top* (OTT).

Os OTT são aplicações que utilizam a Banda Larga Móvel para fornecer serviços aos utilizadores com *Smartphones* ou *Dongles*, alguns dos quais fazendo concorrência directa aos serviços disponibilizados pelos próprios operadores móveis, como é o caso por exemplo do Skype, de *Viber* no mundo do VoIP, ou o *WhatsApp* no segmento dos SMSs. Para salvaguardar o seu *Core Business*, grande parte dos operadores investiu em PCRF's e PCEF's para fazer o policiamento e efectuar o cumprimento de determinadas políticas, com o intuito de controlar a utilização destes serviços sobre a sua rede. A ideia dos operadores móveis nunca será a de sabotar o serviço disponibilizado pelos OTTs, mas é mais no sentido de não permitir que a qualidade destes compita com a qualidade com os seus próprios serviços.

Outro desafio que foi despoletado com os *Smartphones* foi o incremento de Sinalização na rede e o consequente incremento na capacidade necessária para os acomodar com qualidade.

Todos estes investimentos em CAPEX e OPEX efectuados pelos operadores têm que ter um retorno a médio ou longo prazo. Para se conseguir isto os operadores têm que conseguir captar novos clientes e manter os já existentes. Isto consegue-se com boas estratégias de marketing, com preços competitivos mas também disponibilizando a qualidade desejada pelos clientes, para que estes não decidam mudar de fornecedor de serviço de banda larga móvel, e não contribuir assim para os números que mais assustam os *boards* e *chairman's* dos operadores móveis, o *churn*. Portanto, ter forma de aferir a qualidade do serviço disponibilizado na rede do operador torna-se crucial para apresentar resultados, exigir investimentos na rede e defender cotas de mercado.

3.2 Definições

Durante esta dissertação iremos utilizar alguns termos para facilitar a sua leitura e interpretação. Neste capítulo encontra-se um porto seguro para mais rápida referência. Estes são:

Cliente - Subscritor de uma Rede Móvel, o qual tem associado um MS ou um UE e um IMSI.
Core / *Backbone* IP - É uma rede IP que fornece uma solução de transporte flexível escalável e altamente fiável para todos os serviços do operador.
EPC - Rede de core do sistema EPS (LTE+EPC)
Elemento de Rede - É um elemento constituído por Hardware e Software que tem uma ou mais funções específicas na rede GPRS. Exemplo, SGSN, GGSN, RNC, etc,
IP - A camada da rede responsável por endereçar e enviar, entre outros, capotes de TCP através de uma rede.
Interface - Redes interligada que suportam comunicação ponto-a-ponto.
Iu-CP - Trafego de contolo de transporte de entre o SGSN e a rede de rádio 3G/WCDMA.
Iu-UP - Transport o tráfego do utilizador, entre o SGSN e a rede de rádio 3G/WCDMA.
LTE - Rede de radio de uma rede EPS
PDP / Bearer - É um canal de transmissão de informação, com uma capacidade, um atraso e um *bit error rate*, bem definido. Transmite dados em unidades de forma discreta, a que se dá o nome de pacotes.
Packet Core - Rede constituída por um ou mais SGSN, GGSN, MME ou SGW/PGW.
Segmento de Rede - Troço de uma rede que interliga dois elementos da rede GPRS, Por Exemplo, Gn, Iu-UP, Gi, etc.
Sistema de Monitorização - Elemento que filtra o tráfego duplicado das TAP's, de acordo com as configurações aplicadas pelo operador em cada uma das suas portas.
Sonda / Probe - Sistema que simula um EU ou que efectua testes activos na rede. É habitualmente colocado em locais que geram receitas relevantes para o operador móvel, como escritórios, centros das cidades, etc, ou então em partes da rede, para aferir a sua qualidade.
TAP - Elemento passivo colocado nas interfaces entre Elementos de Rede que permite duplicar o tráfego que o atravessa

3.3 Rede GPRS – *General Packet Radio System*

Para entender melhor a solução e o porquê de algumas das decisões tomadas no seu desenho, neste capítulo iremos introduzir alguns conceitos para ajudar nesta clarificação.

O GPRS (*General Packet Radio System*) disponibiliza serviços de dados tanto a redes 2G/GSM (*Global System for Mobile Communications*) como 3G/UMTS/WCDMA (*Universal Mobile Telecommunication System / Wideband Code Division Multiple Access*). Fornece também a solução que permite transportar pacotes de IP (*Internet Protocol*) entre os móveis (*Mobile Stations* – MS) e a Internet, redes corporativas ou a redes de serviço do próprio operador. A intenção deste capítulo é o de dar uma perspectiva global desta tecnologia que irá permitir compreender melhor a solução implementada e que irá ser exposta no capítulo seguinte.

Uma rede GPRS tem como principal objectivo, fornecer os seguintes serviços:

- Um transporte eficiente de pacotes em redes celulares;
- Um uso eficiente dos recursos escassos existentes no rádio;
- Ser um serviço flexível que permita a cobrança dos conteúdos acedidos, baseado tanto no volume ou na duração da sessão;
- Um Estabelecimento célere da sessão de dados e um acesso rápido aos dados desejados;
- Disponibilizar simultaneamente serviços de voz e de dados, entre os quais, *Browsing*, FTP, WAP (*Wireless Application Protocol*) e MMS (*Multimedia Messaging Service*);
- Ligação a PDNs (*Packet Data Networks*) externar recorrendo a pacotes de IP.

No capítulo seguinte iremos descrever de uma forma relativamente sucinta o GPRS num sistema 3G/WCDMA, que é a rede para a qual a nossa solução foi inicialmente pensada e testada. No entanto, esta solução pode perfeitamente ser utilizada tanto em redes mais clássicas como as 2G/GSM, como também nas mais recentes, 4G/LTE. Pode também ser utilizada indiscriminadamente em redes com os mais diversos fornecedores de elementos de rede, pois não depende dos mesmos. Iremos falar tanto da arquitectura como dos elementos de rede envolvidos.

3.3.1 Descrição Geral *End-to-End*

O sistema GSM, o qual já se pode chamar de clássico pois já conta com 20 anos, foi inicialmente desenhado como um sistema de comutação de circuitos. É sobretudo utilizado para tráfego de telefonia móvel entre telemóveis (MS's) ou então entre o telemóvel e as redes de PSTN (*Public Switched Telephone Network*), mas entretanto, evoluiu para também permitir ser utilizado para efectuar transferência de dados de pacotes entre telemóveis ou então entre o telemóvel e a Internet ou então redes corporativas.

O sistema WCDMA é um sistema móvel de terceira geração (3G) que suporta comunicação tanto por comutação de pacotes como por comutação de circuitos. A grande vantagem relativamente ao GSM é o grande incremento nas velocidades de transmissão.

O GPRS é o serviço de dados no qual se efectua a comutação de pacotes e que é comum a ambos os sistemas mencionados anteriormente, GSM e WCDMA, o que os distingue é a interface de rádio utilizada no acesso. No caso do GSM temos entre o elemento de rede de acesso, a BSC, e o elemento de rede de *Core*, o SGSN, a interface Gb. No WCDMA temos entre o elemento de rede de acesso, o RNC, e o SGSN a interface denominada de Iu.

A solução de GPRS para redes WCDMA encontra-se definida nos *standards* (3GPP) no documento 23.060, tanto as interfaces como os elementos de rede.

Na indústria vulgarmente dá-se o nome de *Packet Core* aos elementos da rede de *Core* que servem tanto de âncora para o tráfego que provêm dos elementos de rede de acesso, os RNC's em redes WCDMA, ao qual de chama de SGSN (*Serving GPRS Support Node*), como aos elementos de rede que interligam à Internet, redes corporativas ou redes de serviço, que se chama GGSN (*Gateway GPRS Support Node*) ou mesmo aos elementos de rede fazem o policiamento do tráfego gerado pelos MSs [15].

Nas figuras seguintes indicam-se de uma forma genérica as três arquitecturas de rede definidas pelo 3GPP. De notar que a arquitectura indicada na Figura 11 recorre a uma funcionalidade opcional descrita nos 3GPP 23.006, capítulo 15.6, a que se dá o nome de 3GDT (3G *Direct Tunnel*) mas que nos últimos anos tem vindo a ganhar tracção e interesse entre os operadores móveis, pois permite optimizar a gestão de tráfego na rede como também permite ganhos relevantes de latência, principalmente quando conjugado com funcionalidades que permitem conjugar a localização do RNC onde o tráfego é originado e o GGSN mais próximo deste local.

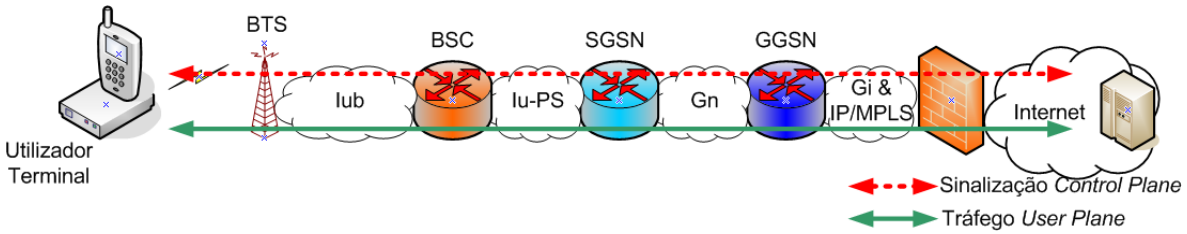

Figura 9 – GPRS, Visão de Alto nível para 2G/GSM

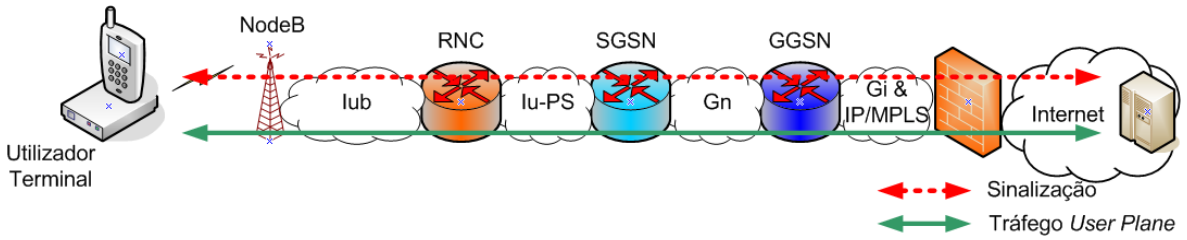

Figura 10 – GPRS, Visão de Alto nível para 3G/WCDMA

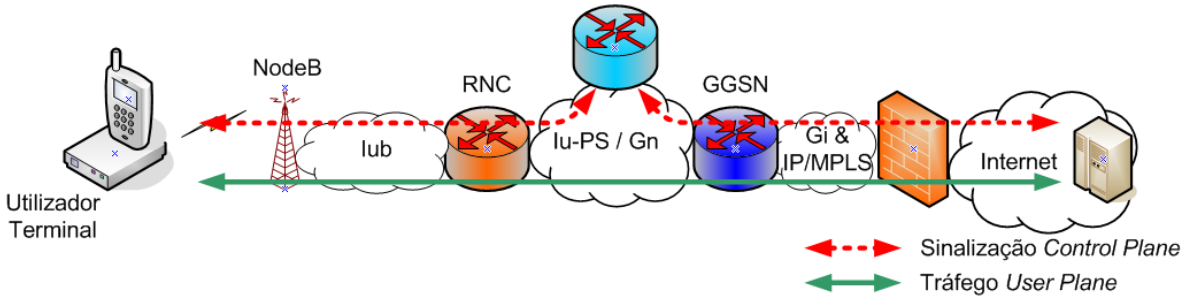

Figura 11 – GPRS, Visão de Alto nível para 3G/WCDMA com 3GDT

De uma perspectiva de IP, que é fundamental para entender o conceito por detrás da solução descrita nesta dissertação, uma rede GPRS pode ser dividida nas seguintes partes:

- A rede de rádio, a qual pode ser considerada como uma rede de acesso na qual um MS se liga a um ou mais PDN's. Os terminais e cliente ligados à rede de acesso, estão acessíveis a outros terminais ou clientes nas PDN's. Ao nível de transporte, as interfaces de redes deste segmento da rede chamam-se Gb, para o caso de GSM, que interliga as BSC's aos SGSN's, e Iu-PS no caso do WCDMA, o qual interliga o RNC ao SGSN, ou GGSN no caso desta rede e APN permitirem 3GDT.
- O PDN, que pode ser a Internet, ou uma rede corporativa ou então uma rede de serviços do próprio operador.
- O *backbone* da rede de GPRS, interliga SGSN's e GGSN's do operador móvel. Pode também interligar os elementos de rede do operador a outros SGSN's e GGSN's de operadores que

tenham um acordo de interligação, a que se chama nos *standards* de *roaming*. A esta interface do *backbone* dá-se o nome de Gn, quando se fala da interligação entre elementos de rede do próprio operador, ou então de Gp quando falamos de interligações de *roaming*.

- Da perspectiva de um MS, o primeiro *Hop* de IP situa-se entre o MS e a saída do GGSN. A segurança definida pelo 3GPP com à utilização dos túneis implementados pela rede de GPRS não permitem ao MS aceder a nenhum elemento ou interface de rede internos da rede de GPRS. Isto é visível na seguinte figura. De realçar, que é devido a esta arquitectura de túneis que as soluções de mercado não conseguem aferir a qualidade dos diferentes parâmetros de rede, entre as diferentes interfaces que recorrem a estes túneis. No entanto a solução apresentada nesta dissertação, foi desenhada para conseguir extrair esta informação.

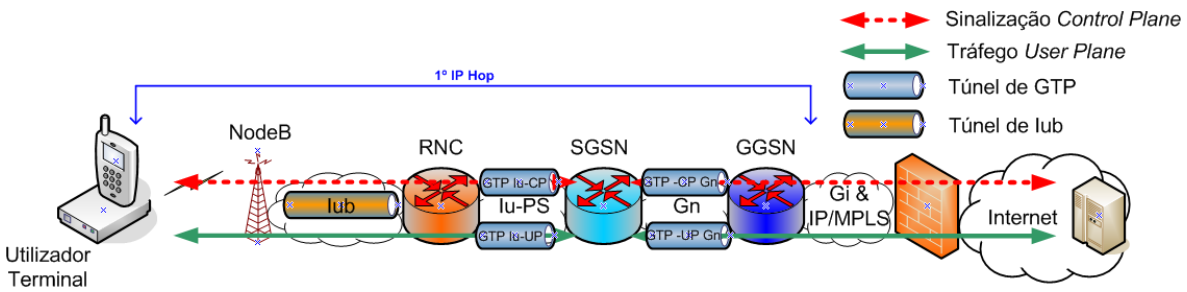

Figura 12 – Segurança/Isolamento na rede GPRS – 1º Salto IP numa rede GPRS

De referir aqui que numa rede móvel de banda larga, o canal ponto-a-ponto (E2E) é constituído por vários componentes independentes uns dos outros, que podem ter características de transmissão distintos [16]. De seguida apresentam-se esses diversos componentes, como definido nos 3GPP.

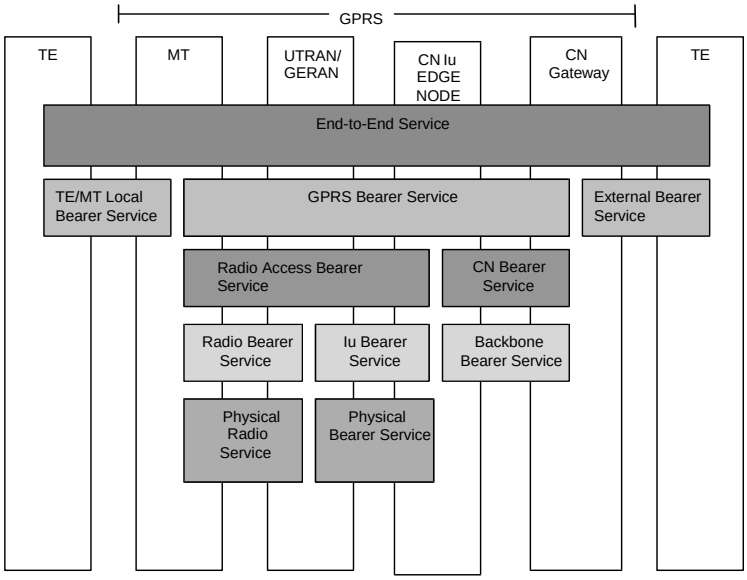

Figura 13 – Arquitectura de QoS E2E

A partir deste momento iremo-nos focar nas redes GPRS que correm sobre 3G/WCDMA. No entanto, vale a pena referir que as diferenças são mínimas, em termos de conceitos, relativamente às redes de GPRS sobre 2G/GSM. Para redes 4G/LTE, alguns destes mesmos conceitos ainda são válidos, mas alguns elementos de rede mudaram de nome, outros desapareceram, noutros foram-lhes acrescentadas funcionalidades.

3.3.2 Arquitectura de Rede

Para se introduzir um pouco mais de familiaridade com as redes 3GPP, é de seguida apresentada uma figura que descreve de uma forma geral todos os elementos e interfaces de rede que existem numa rede GPRS, a partir da *Release* 8 dos 3GPP até à *Release* 10. Os elementos de rede e as suas interfaces irão ser definidos nos capítulos imediatamente a seguir a este.

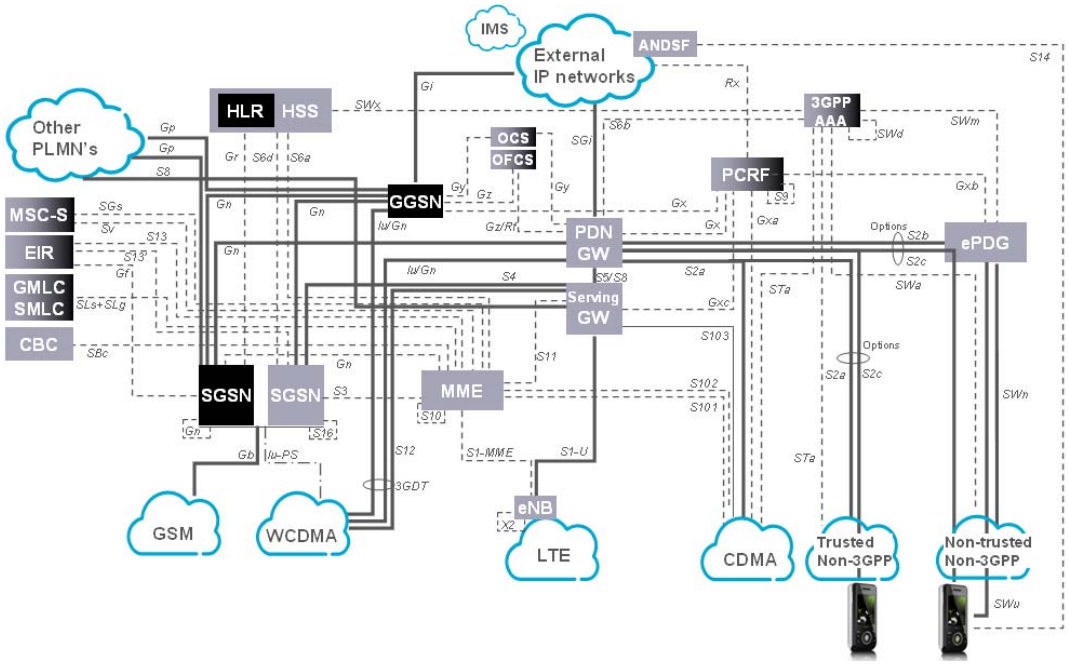

Figura 14 – Descrição da Arquitectura Lógica do GPRS

Os diversos tipos de protocolo de transporte existentes nas mais recentes redes móveis encontram-se ilustrados na figura seguinte. De notar que indicamos uma rede convergente, isto é, em que toda ela a correr tendo como camada de transporte IP sobre *Ethernet*. Este foi o pensamento que se teve no desenho da solução descrita nesta dissertação, em que o tráfego capturado é sempre efectuado em interfaces de rede *Ethernet*. Este princípio está também alinhado com um dos vectores apresentados no capítulo 2.2.

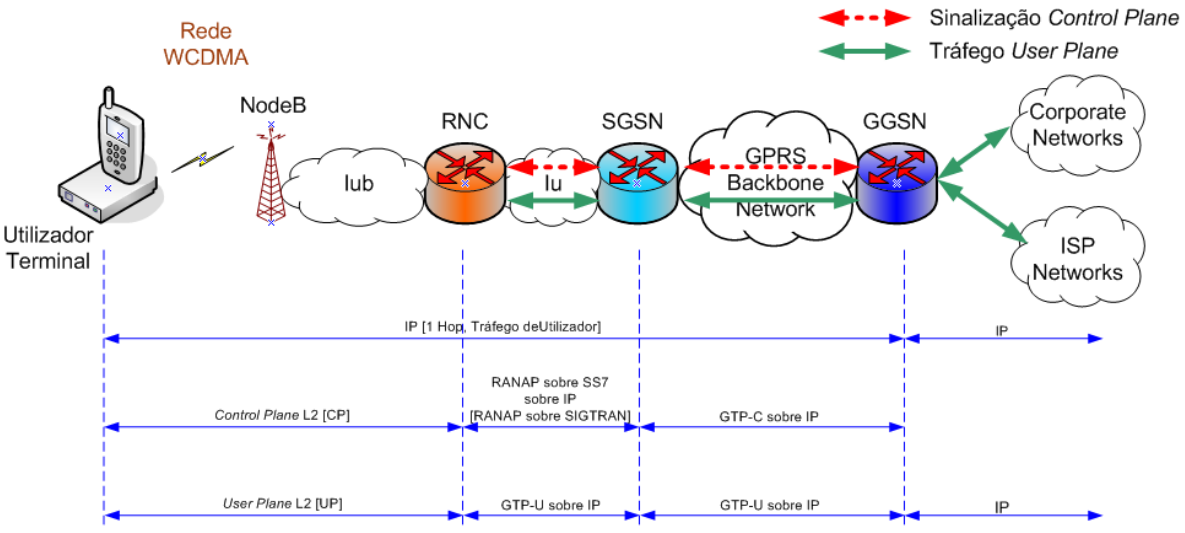

Figura 15 – Tipo de transporte numa rede GPRS, Sinalização (*Control Plane* - CP) e Tráfego do utilizador (*User Plane* - UP)

3.3.3 Elementos de Rede

Da Figura 14 podemos dizer que os dois elementos de rede que são específicos da rede de GPRS são o SGSN e o GGSN, indicados a negro. Os elementos de rede que pertencem ao acesso de rádio diferem se estivermos a considerar um sistema GSM ou WCDMA.

De seguida iremos descrever de uma forma breve o papel de cada um dos elementos de rede relevantes para caracterizar a solução que se irá descrever no capítulo 5, por forma a familiarizar o leitor com as funcionalidades gerais por de trás destes elementos.

3.3.3.1 Mobile Station (MS/UE)

O MS é no essencial uma combinação entre um *Mobile Terminal* (MT) que é um telefone móvel, e um *Terminal Equipment* (TE), que pode ser um computador ligado ao MT. O MT e o TE podem estar também integrados num elemento único, como é o caso por exemplo de um *Smartphone*. O GPRS, no entanto, não cobre esta interface de comunicação, entre o MT e o TE.

Numa rede 3G/WCDMA é no MS onde reside o cartão *User Services Identity Module* (USIM). Este módulo tem um papel essencial numa rede de GPRS pois é nele onde residem as funções de encriptação, a informação da localização e dos serviços GPRS que se pretendem aceder. Este cartão USIM permite identificar o subscritor independentemente do MS utilizado.

3.3.3.2 Terminal Equipment (TE)

O TE contém as aplicações terminais, as quais utilizam a rede de GPRS para transmitir e receber os pacotes de dados. O TE pode ser um *Laptop* ou um *Smartphone*. A rede de GPRS fornece então a conectividade IP entre o TE e o *Internet Service Provider* (ISP), por exemplo. A ligação IP estabelecida entre o TE e a rede de GPRS é estática da perspectiva do TE, isto quer dizer que o TE irá reter o seu endereço de IP até que o MT desligue a sessão de dados estabelecida com a rede de GPRS. Iremos falar nos capítulos seguintes das várias sinalizações trocadas entre o MT e a rede de *Packet Switch* (PS).

3.3.3.3 Mobile Terminal (MT)

O MT é o responsável por estabelecer a comunicação de dados entre o TE e a PDN. O MT encontra-se associado a um subscritor da rede móvel, devido ao USIM que está a utilizar, e

encontra-se ligado pelo acesso da rede de rádio ao SGSN. O MT pode ser comparado a um modem, ligando o TE à parte de PS da rede de GPRS.

3.3.3.4 Radio Network Controller (RNC)

O RNC é responsável por controlar a utilização e integridade dos recursos de rádio. É também ele que efectua as decisões de *handover* na rede, devido à mobilidade inerente a uma rede deste género, e é então responsável pela troca de mensagens de sinalização com o MS. Um RNC pode ser servido por um ou mais SGSN.

Este último caso verifica-se quando na rede se introduz o conceito de SGSN in Pool, o qual permite obter uma maior resiliência na camada dos SGSN.

3.3.3.5 Home Location Registry (HLR)

O HLR é a base de dados na qual está armazenada toda a informação sobre a subscrição dos MS's que têm permissões de aceder à rede de 3G/WCDMA ou 2G/GSM. Nele podem-se encontrar informação sobre, a localização do MS, os vários serviços de dados a que o MS tem permissões de aceder, vulgarmente estes serviços são os chamados APN's (ver mais a baixo), os parâmetros de autenticação, entre outros. De referir que o único elemento de rede da rede de PS que tem contacto com o HLR é o SGSN. A interface utilizada para este contacto tem o nome de Gr, e serve para transferir as várias componentes de informação armazenadas no HLR.

É no HLR onde o operador define o perfil de QoS estático do subscritor, normalizado no 3GPP 23.107 [17], onde se encontram definidas as prioridades no tratamento a dar a um subscritor ou a um grupo de subscritores quando existe concorrência pelos recursos de rede, sendo válido tanto nos recursos de rádio, de transmissão e de *Core*. Entre os mais relevante encontram-se a *Traffic Class*, *Transfer Delay*, *Service Data Unit* e *Maximum Bitrate* de *Downlink* e *Uplink*. Estes parâmetros têm uma influência preponderante na experiência de utilização, mas como foi dito anteriormente, estão também relacionadas como os operadores gerem os recursos e a cadeia de valor da sua rede.

3.3.3.6 Serving GPRS *Support Node* (SGSN)

O SGSN é um nó fulcral na rede de GPRS. É ele que controla a comunicação com os MS's e é também responsável pelo estabelecimento da ligação de dados entre o MS e o PDN. Os vários

túneis que constituem esta ligação encontram-se representadas na Figura 16. Temos então a jusante do SGSN o túnel de GTP (GPRS *Tunneling Protocol*) no interface de Iu-PS entre o SGSN e o RNC, na parte de acesso de rádio e a montante, o túnel também de GTP, mas agora no interface Gn que interliga o SGSN ao GGSN.

A não ser que se utilize a já referida funcionalidade de SGSN em *Pool*, um SGSN serve todos subscritores do sistema WCDMA localizados dentro da mesma área geográfica, habitualmente definida como SGSN *Service Area*. Aos procedimentos de sinalização utilizados pelo SGSN para manter o controlo sobre a localização do MS, dá-se o nome de *Mobility Management*. Nestes procedimentos encontram-se definidas formas que permitem ao MS registar-se na rede, de minimizar interrupções de serviço aquando de mudanças de *Routing Areas* (RA) e até mesmo efectuar *roaming* entre diferentes redes de operadores móveis.

Tem também a tarefa de reencaminhar os pacotes de IP entre os MS's com uma sessão de dados activa e o GGSN. Existe no entanto uma excepção, que é quando em determinado APN se utiliza a funcionalidade denominada 3GDT (3G *Direct Tunnel*), referida anteriormente, que se encontra definidas nos 3GPP 23.060 e 3GPP 23.919. Esta nova arquitectura de rede, que começou a ser estandardizada na *Release* 7 dos 3GPP ficando terminada na *Release* 8, tinha o propósito de separar os elementos de rede onde os túneis de *Control Plane* e de *User Plane* eram entregues. Definiu-se então que o *Control Plane* continuava a ser direccionado tanto do RNC como do GGSN para o SGSN, sendo no entanto os túneis de *User Plane*, que originalmente eram ambos geridos ao SGSN, a serem encaminhados directamente do RNC para o GGSN. Esta funcionalidade tem duas grandes vantagens. A primeira é que o SGSN continuava a ter o controlo da gestão dos recursos e dos túneis e ao mesmo tempo, o de libertar capacidade do SGSN, pois o tráfego de *User Plane* deixa de ter que passar por ele. A segunda, é o que permite optimizar o caminho entre o RNC e o GGSN, diminuindo desta forma a latência na ligação E2E, pois aquando da selecção do APN, o DNS pode ser configurado de forma a seleccionar o GGSN mais perto do RAN, entregando no entanto, sempre pelo menos uma outra alternativa, mais distante, para garantir que continua a haver redundância ao nível da camada dos GGSN's. Podemos dizer que o 3GDT serviu de inspiração para as modernas rede de 4G/LTE/EPC, pois nesta última o elemento de rede que tem o papel idêntico ao SGSN, o MME, já só controla sinalização, todo o tráfego é direccionado entre o eNodeB e a SGW. A figura seguinte representa os túneis, de *Control* e de *User Plane*, num cenário de 3GDT.

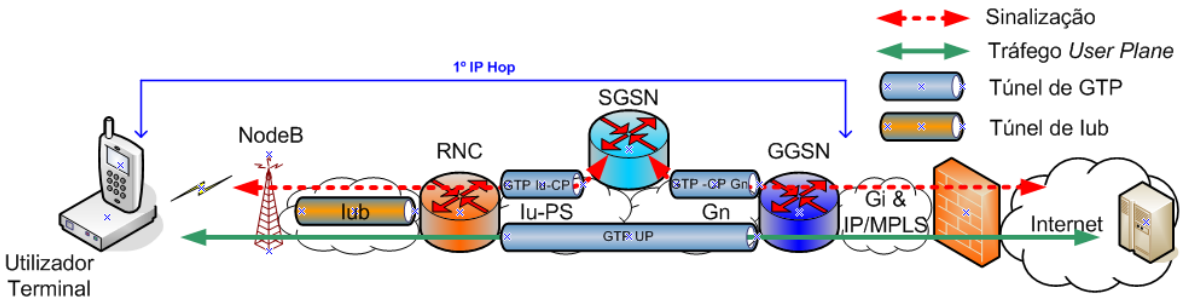

Figura 16 – Apresentação dos túneis de CP e UP num cenário de 3GDT

As ligações entre o SGSN e o MS, e entre o SGSN e o GGSN, são controladas através dos procedimentos de *Session Managements*, entre os quais se encontram activação, modificação e desactivação de PDP *Contexts*. Iremos desenvolver no capítulo 3.4.6 as principais funcionalidades que estão envolvidas nestes procedimentos, que são relevantes para compreender a forma com que os fluxos são conduzidos na rede.

Outra tarefa importante efectuada pelo SGSN é a de fazer a gestão da informação do subscritor. Esta permite ao operador alterar características da subscrição, ou forçar a identificação recorrendo à autenticação do subscritor, garantindo que não é um intruso que está a tentar entrar na rede, ou autorizar ou bloquear ao subscritor o acesso à rede. É também possível diferenciar subscritores e até mesmo serviços, dando a prioridade desejada a cada um deles gerindo assim de uma forma mais eficiente os recursos de rádio colocando a adequada Qualidade de Serviço (QoS). Na rede de Core isto também permite diferenciar o acesso à infra-estrutura da rede IP, diferenciando os pacotes de IP com o DSCP mais apropriado.

Por fim o SGSN efectua, para efeitos de taxação, o registo de cada uma das sessões de dados efectuado por cada um dos MS's. Pode efectuar a recolha do número de *bytes* consumidos durante a sessão, mas somente em sessões que não fizeram uso da funcionalidade 3GDT.

3.3.3.7 Gateway GPRS *Support Node* (GGSN)

O GGSN tem como função primária efectuar o encaminhamento, tanto em *uplink* como em *downlink*, dos pacotes de IP entre o SGSN e o PDN, considerando um cenário clássico de uma rede de GPRS, na qual não existia a funcionalidade de 3GDT. Tem adicionalmente as funções de gerar o registo de taxação para cada uma das sessões dos MS, e efectuar a inspecção do tráfego, em inglês *Deep Packet Inspection* (DPI), dessas mesmas sessões de forma a caracterizar os serviços utilizados, permitindo além de uma taxação diferenciada por serviço, efectuar o cumprimento de regras definidas pelo operador, de forma a possibilitar ofertas de mercado mais atraentes aos

utilizadores. Permite também limitar ou até mesmo bloquear certo tipo de tráfego, como por exemplo, o *Peer-to-Peer* (P2P). Estas regras são habitualmente atribuídas por um elemento de rede externo ao GGSN, chamado PCRF, que se irá descrever no capítulo seguinte.

Em termos de *Session Management*, sucintamente, o GGSN trata da activação, modificação e desactivação dos contextos de PDP, da negociação do QoS a ser atribuído a um determinado contexto de PDP e de distribuir os IP's aos MS's de uma forma fixa ou dinâmica.

3.3.3.8 Policy and Charging Rules Function (PCRF)

A função do PCRF é o de fornecer ao GGSN informação sobre políticas de autorização e de controlo a serem aplicadas à sessão de dados de um determinado utilizador. Desta forma o operador ganha controlo sobre como o tráfego flui na sua rede e é mais um elemento de rede que permite ao operador da rede móvel trabalhar em ofertas comerciais mais apelativas para os seus utilizadores. Denotar que em redes 3G/WCDMA este elemento é opcional, mas nas redes 4G/LTE é um elemento de redes obrigatório, quando se recorrem a *bearer* dedicados, como para voz ou vídeo em tempo real.

3.4 *Session Management* (SM)

Como iremos verificar no capítulo 5 quando descrevermos a solução, o tráfego recolhido irá ser colectado em diversos segmentos da rede e que esses segmentos encontram-se definidos genericamente na norma 3GPP TS 23.060. São nesses segmentos por onde são transmitidas as mensagens de sinalização de *Control Plane* entre os diferentes elementos de rede que fazem parte da rede de GPRS, e por onde fluem os pacotes de tráfego ou *User Plane*, que transportam os dados de e para o MS.

Neste capítulo pretendemos explicar quais os principais fluxos, entre o SGSN e o MS e entre o SGSN e o GGSN, e quais as mais relevantes mensagens trocadas aquando da activação de uma sessão de dados, o conhecido na literatura por activação de um contexto de PDP (do inglês, "*Packet Data Protocol Context*") na rede GPRS e em quais das diversas redes definidas pelas referidas normas, se encontram essas mensagens. Existem outros procedimentos de SM definidos na norma referida anteriormente, mas tais procedimentos não irão ser exploradas, pois não acrescentariam valor à dissertação.

De notar que se irá considerar que o MS terá que estar registado na rede de GPRS. Para tal é necessário que o MS tenha efectuado procedimentos, não de *Session Management* mas sim de

Mobility Management, como as definidas nos 3GPP TS 23.060, capítulo "6 *Mobility Management Functionality*".

3.4.1 Introdução

Os procedimentos de SM servem para estabelecer e para coordenar a sessão de dados entre o MS e a PDN, utilizando como o destino do tráfego um GGSN onde o serviço desejado, ou como é mais vulgarmente conhecido, o APN (*Access Point Name*), se encontra definido e configurado.

Em cada contexto de PDP pode estar associado um TFT (*Traffic Flow Template*), que não é mais do que um filtro com o qual se consegue identificar o tráfego associado a um contexto de PDP específico podendo diferenciar de um outro Contexto de PDP iniciado pelo mesmo MS mas com requisitos de QoS distintos. Desta gestão fazem parte os estados do contexto de PDP. Os estados do Contexto de PDP [28] são indicados pela figura seguinte, e que resumidamente indicam se existem ou não dados a serem transferidos para o endereço de IP associado ao contexto de PDP e ao TFT.

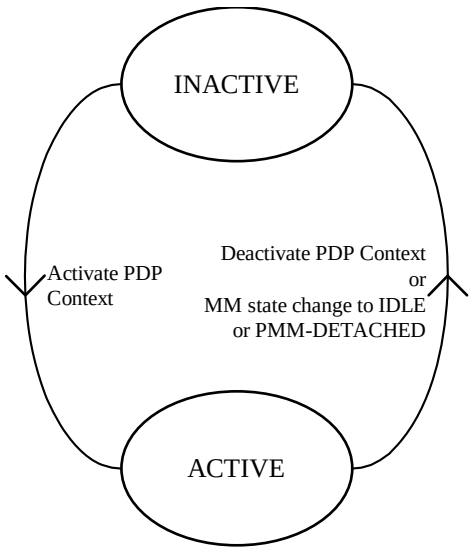

Figura 17 – Modelo de Estados Funcionais do contexto do PDP

Uma sessão de dados é constituída por Contextos de PDP e é um pré-requisito para o MS poder trocar pacotes de dados com o PDN. O PDP é essencialmente um túnel de IP ponto-a-ponto, ente o MS e o GGSN. Cada sessão irá criar três Contextos de PDP em diferentes locais da rede, um no MS, outro no SGSN e o último no GGSN, como de pode verificar na figura seguinte.

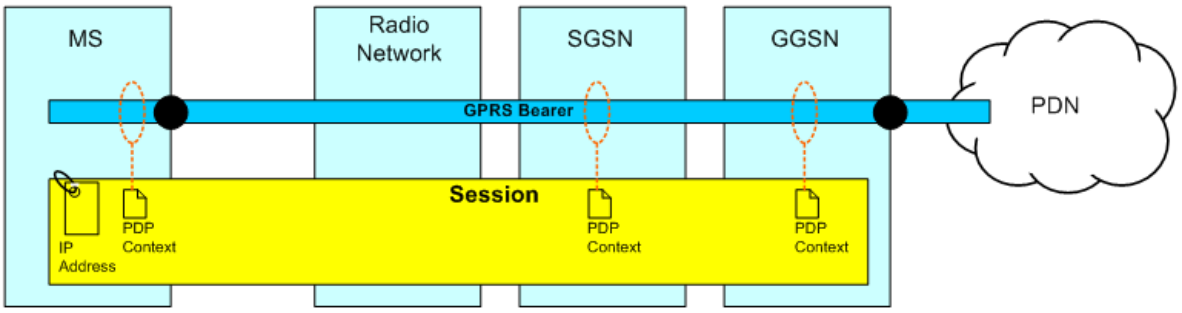

Figura 18 – Visão lógica de uma Sessão de Dados

Um contexto de PDP tem várias funções. Entre as mais relevantes temos a atribuição do endereço IP ao MS e dos diversos parâmetros de Qualidade de Serviço (QoS).

Num sistema 3G/WCDMA o transporte do tráfego do utilizador entre o MS e o SGSN é efectuado através de ligações lógicas a que se dá o nome de RAB (*Radio Access Bearer*), ou mais simplesmente *bearer*, e é este que define o QoS da ligação que irá ser oferecida ao utilizador. Assim diferentes RAB's podem ser disponibilizados dependendo do tipo de subscrições que se encontra associada ao utilizador, e ao tipo de serviço que o MS pretende aceder. Isto é, se o utilizador tem uma subscrição com características mais distintas das ofertas comercias tradicionais, permite-lhe ter alguma diferenciação positiva perante tráfego concorrente. Também permite ter essa diferenciação, se se pretender utilizar uma aplicação que é sensível a atrasos, como por exemplo voz ou *Streaming*, e se se estiver a concorrer com tráfego de Internet como *web browsing* .

3.4.2 *Quality of Service* (QoS)

A possibilidade de diferenciar os utilizadores ou o tipo de tráfego gerado é essencial numa rede móvel de dados moderna. A massificação da utilização dos serviços de dados nas redes móveis e o facto de os recursos de rádio serem cada vez mais um bem escasso e extremamente precioso, como se pode ver na figura seguinte, na qual é dada pela estimativa efectuada pelos analistas da *Analysys Mason* em Março de 2013 depois do leilão de espectro no Reino Unido.

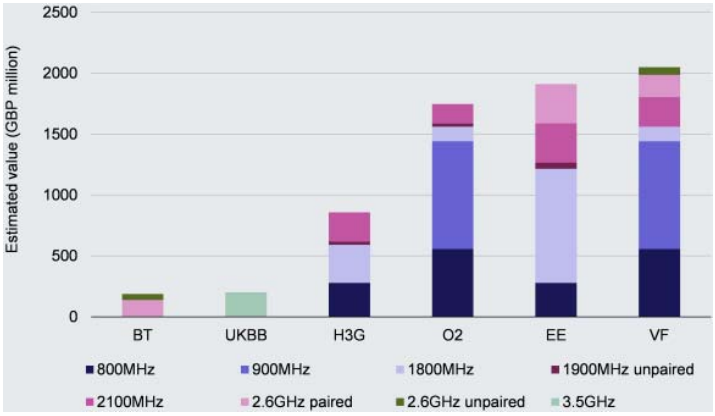

Figura 19 – Estimativas do valor relativo do Espectro por Operador no UK, Março 2013 [18]

Os operadores desejam que os clientes sintam que o serviço prestado pela rede esteja de acordos com as suas necessidades e anseios. Dar a prioridade adequada a cada um dos diferentes tipos de tráfego gerado pelos vários MS e ir ao encontro do que se paga pela respectiva subscrição torna-se uma tarefa obrigatória num operador móvel moderno.

As redes GPRS foram concebidas para permitem ao operador efectuar uma diferenciação muito granular, indo desde o QoS que se atribui ao *bearer* e aos túneis de GTP, aquando da activação do Contexto de PDP baseado na Classe de Tráfego (*Traffic Class*) do utilizador, mas também à importância relativa entre *bearers* na altura da alocação e retenção desses recursos em situação de limitações ou mesmo congestionamento na parte do acesso rádio ou *core*.

Quando o contexto de PDP é inicialmente activado e o MS não explicita qualquer tipo de parâmetro de QoS, o SGSN negoceia o QoS com o valor que existe no perfil do subscritor com o GGSN, com o MS e com o RNC. O QoS pode também ser renegociado quando existe um procedimento de *Inter*-SGSN *Routing Area Update*, isto é, quando o MS tem um PDP activo e passa para a cobertura de um outro SGSN utilizando a mesmo tecnologia de acesso, ou num *Inter-System Change*, que se verifica quando o MS passa para uma outra tecnologia de acesso.

3.4.3 Endereço IP

O endereço IP entregue ao MS aquando da criação do Contexto de PDP pode ser público ou privado, e pode ser atribuído de uma forma dinâmica ou estática. As versões de SGSN mais recentes já disponibilizam a atribuição de endereços, tanto de *Internet Protocol* versão 4 (IPv4) come de *Internet Protocol* versão 6 (IPv6).

3.4.4 Activação de um Contexto de PDP

Esta função de activação, numa rede 3G /WCDMA como definida da norma 3GPP TS 23.060 da *Release* 8, é sempre iniciada pelo MS. De notar que em versões mais recentes do 3GPP já existe a possibilidade de a sessão de dados ser iniciada pela própria rede. Este procedimento é utilizado para estabelecer um canal lógico virtual entre o MS e a PDN, através da rede de GPRS, como indicado na Figura 20. Depois de efectuada a activação de um contexto de PDP com sucesso, o subscritor poderá começar a comunicar com o PDN pretendido, utilizando para tal o GGSN que contém o serviço desejado e que o DNS seleccionou a partir das indicações dadas pelo operador da rede.

Quando o MS requer a activação do contexto PDP, o SGSN efectua uma rotina de controlo de admissão para confirmar que tem recursos internos disponíveis para gerir este novo contexto de PDP. Determina posteriormente qual a QoS a ser oferecida ao MS utilizando para tal a informação da subscrição do perfil recebida do HLR, mas também a informação trocada com o GGSN, com o RNC e até mesmo com o MS, dependendo aqui to tipo de *Traffic Class* indicado pelo MS aquando da activação do contexto de PDP.

A seguir o SGSN irá seleccionar o GGSN onde o serviço requerido pelo MS, definido pelo APN (*Access Point Name*), explicitamente na activação do contexto de PDP ou então definida no perfil. Para tal irá questionar o DNS interno do operador para efectuar a tradução, do APN para o IP do GGSN ou GGSN's. O APN é simplesmente um nome lógico associado a um serviço.

Quando o GGSN estiver seleccionado, dependendo das regras definidas no DNS e também possivelmente por regras internas ao SGSN, irá então entrar em contacto com o GGSN para a alocação do IP ao MS.

3.4.5 Resolução do Nome do APN

É neste ponto que se define qual o GGSN onde irá ficar ancorado o contexto de PDP que se pretende activar.

Um APN é um nome lógico que se refere a um PDN ou a um serviço ao qual o subscritor se deseja ligar. O servidor de DNS (*Domain Name System*) é utilizado para mapear o APN a um endereço de IP, o qual representa o endereço de sinalização do GGSN, também referido na documentação por GTP-CP, ao qual o Contexto de PDP se irá ligar e no qual o serviço desejado se encontra definido.

O SGSN pode utilizar um APN explicitamente requerido pelo MS, ou então um existente no perfil do subscritor. O SGSN pode também ter definido um APN que é utilizado por omissão (*Default* APN). Um ou mais GGSN podem ter definido o serviço explicitado pelo APN, dependendo do nível de redundância que o operador desejar ter nesse serviço na sua rede. Isto quer então dizer que na resposta do DNS podem existir mais do que um IP como resposta à resolução do APN. Desta forma, o que os 3GPP definem é que o SGSN deve começar por iniciar o pedido de activação do contexto de PDP utilizando o primeiro IP da lista de IP's recebida do DNS. O SGSN poderá necessitar, no caso de uma resposta negativa ou mesmo de uma não resposta do primeiro GGSN, avançar na lista de IP's recebido do DNS e tentar efectuar a activação no próximo IP contido nessa lista.

Este ponto é relevante, pois dependendo das políticas e configurações aplicadas ao DNS, como *Views* de DNS, *Round Robin*, etc, podemos ou não optimizar a entrega do tráfego gerado ou direccionado ao utilizador, escolhendo para tal um GGSN co-localizado ou relativamente próximo do SGSN, ou caso contrário, poderemos activar os contexto de PDP para um GGSN numa localização distante, e forçando assim longas distâncias, o que poderá acarretar alguma penalidade na latência sentida pela sessão de dados, e com influência directa na experiência do utilizador.

3.4.6 Fluxos de Tráfego

O nosso interesse no fluxo de Activação de um Contexto de PDP (PDP *Context Activation*) numa rede 3G/WCDMA é o de permitir explicar e caracterizar as mensagens de sinalização trocadas entre os nós da rede GPRS, permitindo ao mesmo tempo explicar quais as redes definidas pelo 3GPP por onde o tráfego do utilizador irá fluir e de onde a nossa solução o irá capturar.

Um operador irá ter à sua disposição diversas alternativas por onde poderá encaminhar o tráfego do utilizador. É necessário que o operador tenha consciência dos seus impactos, tanto a nível da optimização da utilização dos seus recursos de rede, mas também, quais os impactos que as decisões terão a nível de experiência de serviço para o utilizador final.

Nos 3GPP TS 23.060, capítulo "9.2.2.1 PDP *Context Activation Procedure*" encontra-se explicado todo esse fluxo em pormenor. O nosso intuito nos parágrafos seguintes é focar-nos no seu essencial de forma a permitir ter uma ideia abrangente.

Visualmente a activação de um Contexto de PDP é representado pelo seguinte fluxo.

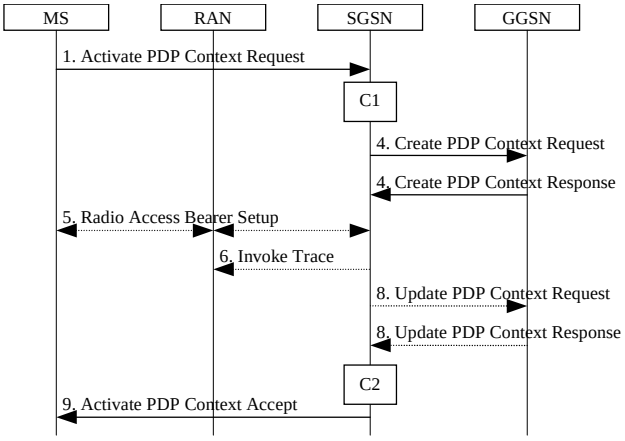

Figura 20 – Procedimento de Activação de um Contexto de PDP num sistema WCDMA

Inicialmente o MS envia uma mensagem de pedido de Activação de Contexto de PDP "*Activate* PDP *Context Request*" para o SGSN. O MS deverá, se o desejar, explicitar o QoS ambicionado, se pretende ou não um IP especifico e qual o APN a que se pretende ligar. O SGSN irá validar toda esta informação com o perfil que se encontra definido no HLR e que já se encontra na posse do SGSN.

Se o pedido for válido, o SGSN resolve o APN no DNS Interno, e irá receber deste uma lista com os GGSN's que têm configurado o serviço pretendido pelo MS. O SGSN irá enviar o pedido de "*Create* PDP *Context Request*" para o primeiro GGSN da lista recebido do DNS. É nesta mensagem que o SGSN informa o GGSN sobre os IP's do seu lado que irão ser utilizados para as mensagens de sinalização e para o fluxo de tráfego.

Se o GGSN aceitar o pedido recebido do SGSN irá enviar uma mensagem "*Create* PDP *Context Response*" para o SGSN onde irá também indicar os seus IP's de sinalização e de tráfego.

Em 3G o SGSN irá trocar informação com o RNC de forma a activar o RAB entre este último e o MS, que irá ser responsável por transportar o tráfego entre o MS e o SGSN.

Se a activação do RAB for efectuada como sucesso, então o SGSN irá enviar a mensagem "*Activate* PDP *Context Accept*" para o MS.

No caso de o SGSN permitir activações de 3GDT para o APN pedido pelo MS, iria haver uma mensagem adicional para o GGSN, na qual o SGSN modifica o IP que será utilizado para tráfego, alterando o seu IP para o IP fornecido pelo RNC aquando da criação do RAB. A referida mensagem será uma "*Update* PDP *Context Request*" e será enviada para o GGSN, depois da activação do RAB. É observando estas mensagens e os IP's transferidos que conseguimos identificar o trajecto que o fluxo de tráfego de um utilizador irá seguir, e desta forma, identificar quais os pontos de monitorização da nossa solução que irão ser utilizados para capturar este mesmo tráfego.

4. Experiência do Utilizador

Neste capítulo iremos falar sobre a importância dos operadores móveis conseguirem obter informação sobre alguns dos KPI de rede relacionados com a qualidade de rede, e como se pode tratar essa informação de forma a fornecer indicações sobre como classificar a experiência de utilização dos serviços disponibilizados.

4.1 Introdução

Com a saturação do mercado de banda larga móvel, a intensa concorrência neste mercado [19] e as regulamentações legais que permitem a portabilidade do número de telefone, facilitando assim a mudança do prestador de serviço, torna-se extraordinariamente importante para o operador móvel medir todas as interacções que os seus cliente têm com a organização, tanto a nível humano como técnico, e qual a qualidade ou percepção da mesma. Desta forma consegue-se ter uma medida sobre a fidelização dos clientes, permitindo aferir qual a potencial percentagem que o cliente que poderá estar a ponderar mudar para um outro operador. Em Inglês dá-se o nome de *churn* a esta migração entre operadores. Torna-se assim crítico existirem processos que permitam, tanto aferir a satisfação quando um subscritor interage com o serviço de apoio ao cliente, por exemplo, aferindo em quanto tempo o seu problema demorou a ser resolvido mas também a satisfação sobre a forma como foi atendido, como também é relevante aferir todos os outros factores chave para o subscritor, como os que são indicados em [2], que vão deste os já referidos factores humanos com o apoio a cliente, mas também técnicos como a qualidade da voz e a qualidade do serviço de banda larga móvel, passando também pelas questões monetárias como a taxação dos diversos serviços.

O nosso interesse incide nas características técnicas da interacção do subscritor com a organização, mais precisamente com a qualidade da rede de banda larga móvel do operador. Iremos explicar como este último poderá enriquecer o conhecimento sobre a satisfação dos seus clientes nas diversas interacções com os serviços disponibilizados.

4.2 Parâmetros de Rede

Na literatura existem diversas fórmulas que expressam e correlacionam alguns dos parâmetros que caracterizam a rede com a experiência de serviço. A que inspirou esta dissertação

[20] permite obter boas aproximações com a realidade e relaciona o *Throughput* máximo oferecido pela camada TCP com alguns parâmetros influentes e que caracterizam a transmissão E2E como o *Round-Trip Time* (RTT), que está relacionada com a latência dos pacotes em ambos os sentidos, *downlink* e *uplink*, e por fim com o *Packet Loss*.

$$Throughput_{Maximum} = \frac{MSS}{RTT_{\min} * \sqrt{p}}$$

Fórmula 1 - *Throughput* Máximo da camada TCP oferecido por uma rede

O significado dos diversos parâmetros desta fórmula estão expostos de seguida.

- **MSS**

- *Maximum Segment Size* (Tamanho máximo do datagrama).
- É habitualmente definida pelo sistema operativo. No entanto existem alguns fornecedores de SGSN, GGSN ou SGW/PGW que permitem alterar este parâmetro, na altura da sua negociação, para evitar fragmentação devido ao encapsulamento de GTP utilizada em algumas das suas redes.
- Este parâmetro é enviado por cada uma das extremidades da ligação aquando do TCP *3-Way Handshake*, no SYN e SYN-ACK, respectivamente.

- **RTT_{min}**

- *Round Trip Time* Mínimo. É um parâmetro que é dependente da arquitectura de rede, da transmissão utilizada e do próprio equipamento tanto no acesso, por exemplo, BTS, NodeB e eNodeB, BSC e RNC, e mesmo do elementos de *Core*, como SGSN, GGSN, PGW e PCEF.
- Está intimamente ligado à latência disponibilizada pela rede tanto no *downlink* como *uplink*. A latência irá ser descrita com maior detalhe no capítulo 4.4 “Lâtencia e Round Trip Time (RTT)”.
- Este último parâmetro é o que pretendemos compreender melhor com a solução que se desenvolveu e que irá ser exposta nesta dissertação, no entanto iremos além de calcular o RTT, dividi-lo nas suas duas componente, tanto de *donwlink* como de *uplink*, pois a solução permite acrescentar essas duas dimensões e que a diferencia das outras soluções do mercado.
- De referir que temos interesse em determinar o RTT_{min} pois é o indicador do valor mínimo que a rede móvel em questão consegue alcançar. De notar que na determinação do RTT_{min} não se inclui a parte de sinalização de TCP como o SYN e o FIN *Handshake*, pois o tamanho dos pacotes

envolvidos não são representativos de uma transferência de dados real, pois são utilizados pacotes com tamanhos inferiores a 100 *bytes*.

- É de referir que nem todos os valores de RTT durante a duração da vida da transmissão têm que corresponder ao valor mínimo. Isto é referido na documentação dos 3GPP [21]. Isto deve-se a que a camada de TCP, quando a transferência de dados está a decorrer e não se encontram questões com a TCP *Congestion Window*, os TCP ACK's enviados pelo cliente para o servidor não são feitos pacote-a-pacote, isto é, a um pacote enviado pelo servidor resultaria em um pacote de ACK enviado pelo cliente para o servidor, mas sim esses TCP ACK's enviados representam o reconhecimento que o cliente recebeu um conjunto de pacotes enviados pelo servidor. Isto é visível na Figura 24 e é por isso que o RTT em velocidade de "cruzeiro" da transmissão pode chegar a valores bastante superiores ao RTT_{min} sem que por isso a velocidade de transmissão tenha qualquer degradação.

- ***p***

- *Packet Loss*. Depende das condições de rádio e dos próprios elementos de acesso, transporte e de *Core*.
- Especifica o número de pacotes que são perdidos na rede durante a transmissão. Esta perda pode ser devida a vários factores como, a corrupção do pacote no meio de transmissão ou perdido num ponto da rede em que haja congestão devido à falta de espaço nas filas de espera, em inglês *buffers*, tanto nos interfaces de entrada como de saída dos elementos de rede.
- Numa moderna rede de transporte e de Core estes eventos devem ser muito raros, se a rede estiver bem dimensionada e desenhada para o universo de subscritores e para o perfil de tráfego dos seus clientes. Dito isto, como valor de referência para este parâmetro poderá ser da ordem dos 10^{-4}.

4.3 Impacto na Experiência do Utilizador

Destes dois últimos parâmetros, o *packet loss* é o único que é possível recolher valores estatísticos dos vários elementos de rede. Este valor pode ser representado por pacotes descartados devido à sobrecarga da capacidade nas filas de espera, ou devido a mecanismos de protecção de sobrecarga dos processadores. Por sua vez, para obter RTT o operador não tem forma directa de aferir o seu valor estatístico dos elementos de rede, nem de uma forma E2E, nem por elemento ou segmento de rede. Para se aferir o seu valor, tem que se recorrer a testes

activos, podendo estes mostrar o valor de RTT de uma perspectiva E2E, ou então, como é o caso da solução criada, evidenciar a latência em cada um dos sentidos de uma forma E2E, mas também a contribuição para esta latência E2E que cada segmento de rede introduz no valor final.

Como foi dito anteriormente o RTT é composto pela latência sentida pelo tráfego do utilizador tanto no sentido de *downlink*, entre o Servidor e o UE, como no sentido oposto de *uplink*, entre o UE e o Servidor. É mais relevante em ligações de dados sobre TCP. Existe no entanto tanto na literatura como nos próprios operadores móveis, propostas de aferições de medidas de latência da redes que se baseiam em tráfego ICMP que corre sobre UDP, em que o método mais amplamente utilizado é recorrendo a PING's, devido principalmente à facilidade de obtenção de valores RTT, mas utilizando para tal, tamanhos de pacotes distintos dos que caracterizam os serviços habitualmente utilizados pelos utilizadores. Iremos no capítulo 4.4 argumentar e mostrar a nossa visão sobre as desvantagens deste método para a aferição do RTT numa rede móvel de banda larga.

Analisando a literatura e as diversas perspectivas de como definir um *Key Quality Indicator* (KQI) que enquadre os diversos indicadores recolhidos da rede num indicador único de satisfação do cliente ou também referido como experiência do utilizador, em inglês *Customer Experience*, podemos concluir que quantos mais dados o operador tiver à sua disposição, de diferentes origens, e que esses dados possam ser utilizados de forma a alimentar o algoritmo que consiga fornecer um indicador de experiência do utilizador, maior precisão e logo melhor informação terá à sua disposição para o ajudar a decidir como evoluir, investir na sua rede e reter os seus clientes.

Com a solução que se irá expor nesta dissertação, pretende-se dar um passo mais à frente, permitindo ao operador utilizar mais uma variável que lhe permita aferir em tempo real a qualidade da experiência do utilizador, deduzida pelo tráfego gerado por utilizadores reais da sua rede e não por ferramentas que tentam simular e aferir de uma forma indirecta a qualidade dessa experiência.

Observando o que o TMForum recomenda, em como aferir a experiência do utilizador, temos o seguinte diagrama.

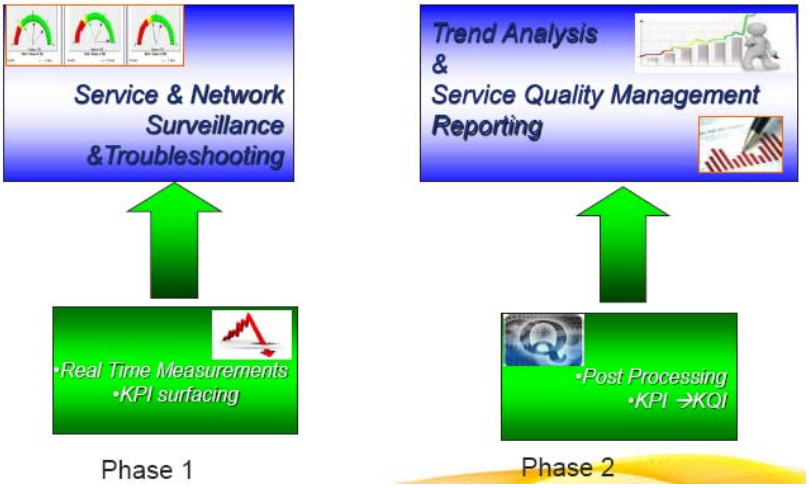

Figura 21 – Áreas Estratégicas de Negócio

A solução proposta irá permitir então recolher mais e melhores dados do que um operador consegue actualmente recolher na sua rede, e irá então permitir definir com maior granularidade KPI's que podem ser tratados e inferidos em KQI's mais ricos, colocando-os ao dispor do operador. Para enquadrar na figura anterior, conseguir-se ter mais um fluxo de KPI's de rede, indicado na figura como "*Phase* 1", que está intimamente ligado ao tráfego que o utilizador está a gerar, consegue-se acrescentar valor aquando do cálculo dos KQI's, na figura indicado como "*Phase* 2", possibilitando enriquecer os relatórios gerados e que dão a visibilidade necessária sobre a qualidade do serviço sentido cliente final.

Estes KPI's e KQI's têm a grande vantagem de permitirem ao operador ter um painel de instrumentos operacional da sua rede que o pode ajudar a identificar com rapidez e precisão, tanto de uma forma E2E como por troços de rede ou então pelos elementos da sua rede, que são responsáveis pela degradação do serviço prestado.

Para ir ao encontro do que um cliente ambiciona de uma rede móvel de banda larga, que é essencialmente uma boa experiência de serviço logo, acesso célere ao serviço e um *throughput* constante e elevado, os valores de latência e RTT têm de estar dentro de intervalos de valores bem definidos. Os valores teóricos, obtidos pela Fórmula 1, são indicados no gráfico seguinte.

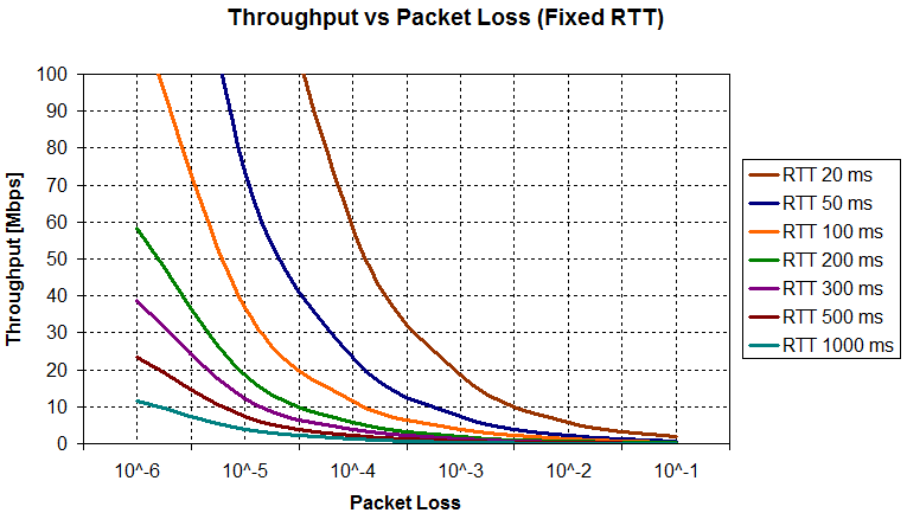

Figura 22 – Valores Teóricos *Throughput* vs. *Packet Loss*, com RTT fixo

A solução apresentada nesta dissertação permite ao operador móvel posicionar a sua rede, ou as várias sessões de dados dos seus utilizadores, sobre estas linhas alertando-o para degradações que possam estar a ocorrer na sua rede, indicando com precisão onde essas degradações se estão a verificar e mostrando troços da sua rede onde possíveis melhorias de rede poderão ser introduzidas, recorrendo para tal a optimizações, a novos equipamentos, a versões mais recentes de *software* ou chegando ao ponto de modificar o desenho da sua rede.

4.4 Lâtencia e *Round Trip Time* (RTT)

De seguida iremos definir dois tipos de latência e *Round Trip Time* (RTT) que existem em redes móveis, e iremos também explicar o impacto que estes parâmetros têm no comportamento das aplicações que correm sobre TCP e UDP.

4.4.1 Definição de Latência

A latência de um pacote de dados, que de vez em quando também é referida na literatura como o "atraso" do pacote, é habitualmente definida como o tempo que demora esse pacote a ser recebido no ponto remoto da rede, depois de ter sido transmitido do seu ponto de origem [21]. O seu valor é influenciado pelo número de elementos de rede que o pacote tem que atravessar e pela distância entre cada um dos elementos de rede.

No essencial existem 3 factores que determinam o atraso que o pacote sofre em cada elemento de rede atravessado. Estes encontram-se definidos de seguida.

- **"Serialização" ou Atraso da Transmissão**

É o tempo que demora a um dispositivo enviar o pacote à velocidade de transmissão de saída dessa interface. As duas dependências para este atraso são o tamanho do pacote de saída e a largura de banda da interface.

- **Atraso de Propagação**

É definido como o tempo que demora um bit a ser transmitido por um elemento de rede, e ser recebido por um outro elemento. É intimamente dependente do meio de transmissão e da distância, sendo independente da largura de banda do segmento de rede.

- **Atraso de Comutação**

Pode ser definido como o tempo que demora a um dispositivo de rede iniciar a transmissão de um pacote depois de o ter recebido. Depende essencialmente, da carga de esforço incutida a esse elemento de rede, logo do número de pacotes em trânsito neste ponto da rede.

4.4.2 Tipos de Latência em Redes Móveis

Nas redes móveis 3GPP existem dois tipos de Latência, uma relacionada com o *Control-Plane* e outra relacionada com *User-Plane*. Nos dois sub-capítulos que se seguem iremos lançar alguma luz sobre elas.

4.4.2.1 Latências de *Control-Plane*

A latência relacionada com o *Control-Plane* está associada à sinalização ou *call setup*, isto é, ao intervalo de tempo sentido pelo terminal do cliente quando se transita entre os estados do RRC (*Radio Resource Control*), definidos pelo 3GPP TS 25.331 [22], desde o estado passivo ou *idle* para um estado em que o equipamento pode receber e enviar dados. Os dois estados nos quais o terminal pode enviar dados têm o nome de FACH ou DCH, tendo cada um deles características distintas em termos de latência, logo em termos de experiência de serviço. Nesta dissertação não iremos aferir esta latência, podendo no entanto a solução desenvolvida ser utilizada em medições deste género.

A forma como o 3GPP definem as transições para o estado activo depende da quantidade de dados que têm que ser transmitidos num determinado instante. O racional por detrás disto é o de conservar bateria do terminal e os recursos de rede. Em [23] é indicado qual o comportamento da camada RLC quando existem dados a serem enviados, tanto em *uplink* como em *downlink*, indicando os valores de referência, que o UE tem que preencher em termos que dados a transmitir, em *bytes*, que permitem saltar para estados onde a velocidade de transmissão é mais generosa.

Este tipo de latência, em sistemas 3GPP 3G/WCDMA HSPA, é caracterizada pela figura seguinte, ao centro a azul.

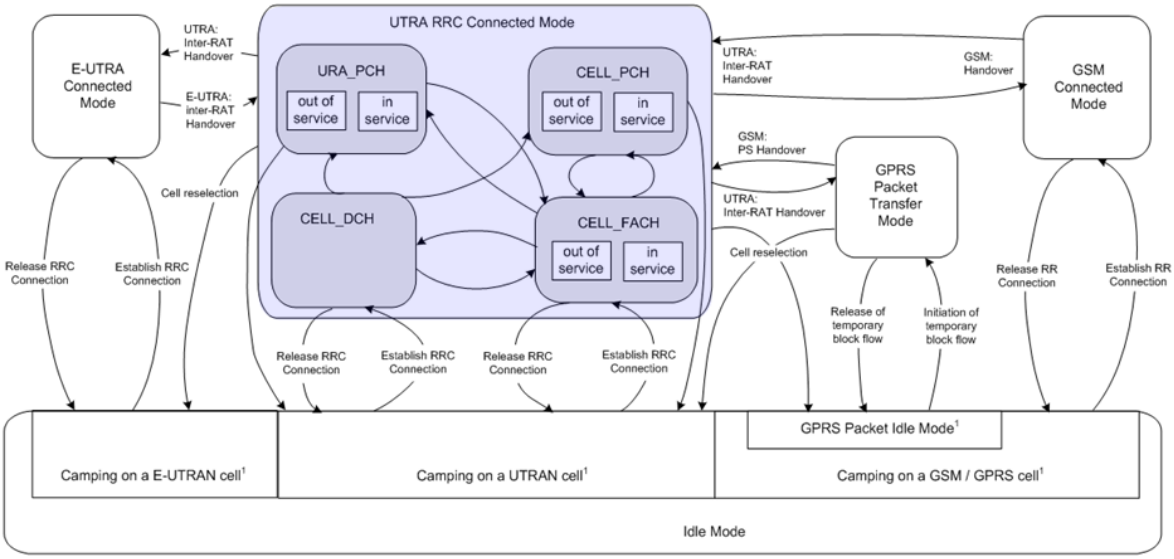

Figura 23 – Estados do RRC e Transacções de Estado em GSM e E-UTRAN (WCDMA/HSPA) [22]

Nas redes móveis de banda larga modernas são então utilizados diversos canais de rádio, com características distintas de forma a melhor gerir os diferentes estados dos utilizadores e as necessidades de transmissão em cada um desses estados. Os três principais estados encontram-se descritos a seguir. De notar, no entanto, que o intuito deste parágrafo não é efectuar uma exaustiva descrição das características de rádio numa rede móvel.

Os estados mais relevantes existentes numa rede 3G/WCDMA são descritos de seguida.

- **Modo *Idle***

Neste estado não existe uma ligação lógica entre o MS e a rede de rádio, logo não existe uma ligação ponto-a-ponto, entre pares de entidades de RRC's (do Inglês, *Radio Resource Control*), estabelecida entre os referidos elementos. Desta forma a rede minimiza a alocação de recursos pelos terminais.

- **Modo CELL FACH (*Forward Access Channel*)**

Neste estado o terminal encontra-se com uma ligação à rede que os 3GPP definem como *Connect Mode* e que por conseguinte existe uma ligação ponto-a-ponto de RRC entre o MS e a rede de rádio. Aqui o MS utiliza os canais comuns de transporte denominados de RACH e FACH, os quais são partilhados pelo universo de terminais abrangidos pela cobertura da antena de radio. Este canal é caracterizado, devido a ser partilhado, por ter uma latência elevada, pois este estado não é utilizado com o intuito de efectuar transferências de dados, é somente um porto de abrigo para a frota de MS quando estão à espera que o utilizador gere ou receba tráfego.

- **Modo CELL DCH (*Dedicated Channel*)**

Como no modo anterior, aqui também existe uma ligação à rede de rádio. Para um MS se encontrar neste modo, o utilizador desse terminal tem que se encontrar a efectuar alguma transferência de dados, no sentido *downlink* ou *uplink*. Aqui o terminal encontra-se a utilizar um canal dedicado, tanto para tráfego como para sinalização. A característica essencial deste canal é o de possuir a latência mais reduzida que se pode ter em qualquer canal disponibilizado em redes 3G/WCDMA.

A forma como um subscritor comuta entre estados ou seja entre canais de rádio é configurável pelo operador. Como os recursos de rádio não são ilimitados, o operador tem que optimizar a utilização dos recursos disponíveis e a satisfação dos clientes no que respeita à experiência de utilização. Assim, as redes permitem utilizar, por exemplo, o ritmo de transmissão em bits ou pacotes por segundo, para decidirem se um subscritor deverá saltar de *CELL FACH* para *CELL DCH*, isto é, de um canal com elevada latência e de recursos partilhados para um com uma latência reduzida e com recursos dedicados.

Habitualmente os clientes encontram-se em *CELL FACH*, onde a latência é mais elevada e onde não existem recursos da redes rádio dedicados à sua ligação, pois grande parte do tempo de uma ligação de dados numa rede móvel de banda larga não existem dados a serem transferidos entre o MS e a rede de rádio. São nestes canais que habitualmente um cliente se encontra quando começa a efectuar tráfego de *User-Plane*, logo a sua experiência de serviço inicial pode ser afectada. Os 3GPP's definem que um cliente salte para *CELL DCH*, onde a latência é muito mais baixa que em *CELL FACH* e onde o cliente tem uma melhor percepção de serviço, quando existe uma demanda de tráfego considerável.

4.4.2.2 Latência de *User-Plane*

O segundo tipo de latência está associado ao *user-plane*, e é a que se verifica a nível aplicacional, quando se trocam dados entre o terminal e a rede. É nesta última que a solução desenvolvida irá incidir.

O seu valor irá ser caracterizado pelas diversas componentes já definidas no capítulo 4.4.1, no entanto simplificando o conceito por detrás do RTT existente em ligações que correm sobre TCP, sendo este último representado pelo somatório da latência no *downlink* sentida por um ou vários pacotes enviados pelo servidor para o cliente, com o somatório da latência em *uplink* sentida pelo pacote de *acknowledge* de um ou mais pacotes enviados, incluído também o

processamento por parte do cliente na extremidade receptora dos pacotes [24]. Isto é visível no seguinte diagrama.

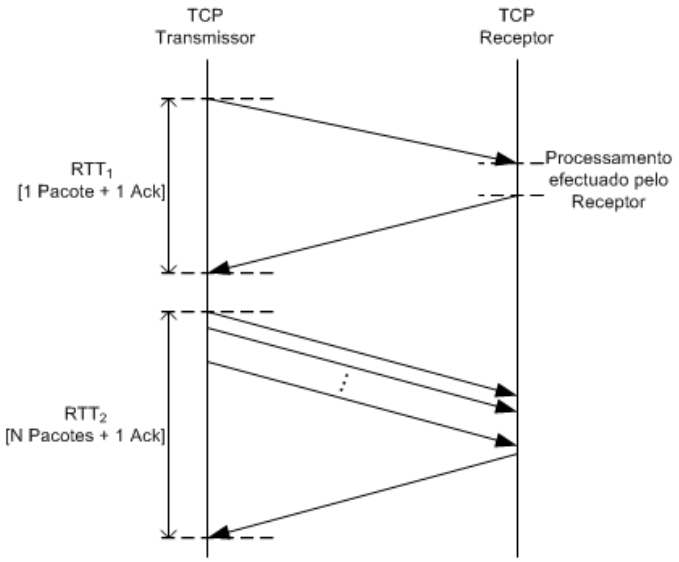

Figura 24 – Medição do *Round-Trip Time* (RTT) E2E

A latência E2E de um pacote num determinado fluxo de tráfego é então a soma de todos os atrasos indicados acima, sofrida em cada um dos saltos, em inglês *hops*. O atraso sentido pelos pacotes neste fluxo não são no entanto constantes em cada segmento da rede, pois nem todos os pacotes enviados no mesmo sentido serão do mesmo tamanho, nem a utilização dos recursos dos elementos de rede serão constantes, como é o exemplo do *Central Processing Unit* (CPU), nem por fim, a ocupação das diversas filas de espera de entrada ou de saída atravessadas dos diversos elementos de rede terão sempre o mesmo volume. A este último atraso habitualmente dá-se o nome de *jitter*.

4.5 Impacto da Latência

O impacto que a latência tem na percepção da qualidade de serviço é enorme, como já tinha sido mostrado no capítulo 4.3, pois tem uma influência directa no *throughput* máximo que a rede consegue disponibilizar de uma forma E2E, mas como se viu no capítulo 4.4.2, também tem influência na velocidade a que um utilizador consegue aceder a um serviço, isto é quando lhe é permitido ter ligação à rede.

Fazendo uma pequena analogia com o mundo das tecnologias de informação, esta última latência está relacionada com a velocidade a que um computador se liga, permitindo ao utilizador abrir o seu *browser* para ver uma qualquer página na Internet. Por sua vez, a primeira latência

tem uma relação directa com a velocidade que o conteúdo da página da Internet é carregado no *browser*.

4.6 Utilização do PING para Aferir Latências em Redes Móveis

Um dos KPI's de rede que habitualmente os operadores móveis são comparados, tanto com os outros operadores nacionais como internacionais, é a latência *end-to-end*. Por comodidade e simplicidade a quase totalidade destas medições são efectuadas recorrendo a pedidos de ICMP, vulgarmente conhecidos por PING's, enviados de um UE para um servidor na Internet, esperando então pela resposta do mesmo. Este intervalo de tempo é então calculado pelo computador ou pela aplicação que está a correr num *Smartphone*, que iniciou os pedidos de PING's.

Este valor será sempre um valor que não representa o RTT efectivamente sentido por um cliente quando gera tráfego com o seu *browser* ou aplicação, pois o perfil de tráfego efectuado tem características completamente distintas do tráfego gerado pelos pedidos de ICMP. Os PINGs recorrem a pacotes com tamanhos, tanto em *downlink* e *uplink*, que rondam os 32 bytes, têm um ritmo reduzido de envio de pacotes, e finalmente por terem uma cadência síncrona.

De referir que o PING foi desenvolvido como ferramenta de teste para conectividade, no entanto é utilizado amplamente em *benchmarks* na aferição do valor de RTT E2E, mais por comodidade do que por garantia de precisão das medições, com o intuito de medir a possível qualidade de experiência do utilizador. Em [25] demonstra-se que em redes móveis, para se alcançar uma precisão satisfatória em medições de atraso na camada de IP, deve-se recorrer a um elevado número de amostras e que essas amostras sejam, tanto em *uplink* como em *downlink*, e que estas sejam aleatórias entre elas. Todos estes requisitos não são possíveis de serem satisfeitos recorrendo à utilização de PINGs, no entanto o método proposto nesta dissertação satisfaz todos estes requisitos.

Como o PING é uma medição RTT, não é possível determinar o seu valor em cada um dos sentidos, isto é, não permite ter a visibilidade do atraso do pacote nos dois sentidos, *downlink* e *uplink*. Como é indicado em [26], existem exemplos na literatura onde se tenta assumir que os atrasos em ambos os sentidos são simétricos, o que não é de todo verdade nas redes móveis. Isto indica que se se quiser ter a visibilidade total da rede e do seu comportamento nos dois sentidos, recorrer a PING's não é solução, é necessário algo mais robusto.

A figura seguinte ilustra os estados que o UE tem que passar ao longo do tempo, quando se utilizam PINGs com tamanhos de pacotes normais. É notório que ao início existe a passagem do

estado CELL_PCH ou *Idle* em que não é possível enviar tráfego para a rede, dá-se então o primeiro salto para o estado CELL_FACH, no qual já é possível enviar tráfego, mas como é um canal partilhado, a potência de rádio não é a maior e o *throughput* que é possível obter também não é o mais interessante, dai o RTT obtido pelos PING's ser mais elevado neste estado do que os RTT's que se conseguem obter no estado imediatamente a seguir. Este último estado é o denominado CELL_DCH, no qual o UE já tem orientação da rede para utilizar mais potência e utilizar um canal dedicado para os dados que tem para transmitir, no qual se conseguem obter *throughputs* mais elevados, que provém do facto de se terem os RTT's mais reduzidos. Assim, dependendo da metodologia utilizada nos testes de *benchmarks*, os valores de RTT obtidos podem ser distantes ou relativamente próximos de uma realidade, se se conseguir garantir que o MS esteja permanentemente em CELL DCH. Mas estarão sempre longe do que o subscritor efectivamente sente e que afecta a sua qualidade da experiência de serviço.

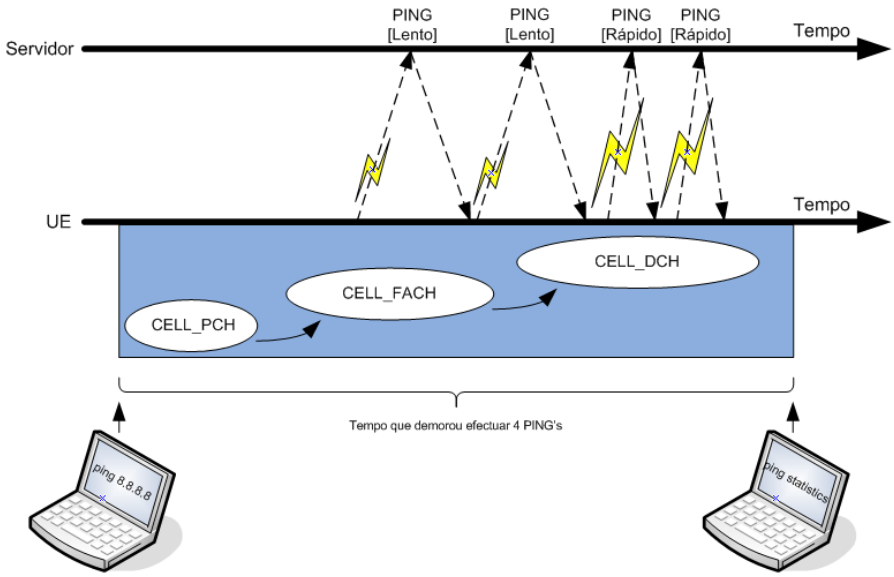

Figura 25 – Evolução dos Estados de RRC ao longo do tempo quando se testam PING's

4.7 Estatísticas de Rede

Um operador móvel tem à sua disposição diversas possibilidades de aferir determinados parâmetros que influenciam a qualidade da rede. Entre os mais utilizados temos os seguintes.

- **Estatísticas dos vários elementos de rede**

Elementos de rede com os nodeB's, RNC's, SGSN's e GGSN's, para 3G ou eNodeB's, MME, SGW's e PGW's, para 4G, ou então elementos da infra-estrutura de IP como *routers* em *switches*, que serão comuns às duas redes. Aqui podemos encontrar estatísticas como por exemplo, perda de pacotes, tanto na parte de *Control Plane* como na parte de *User Plane*, em *uplink* ou em *downlink*.

- ***Service Quality Management***

Estas soluções permitem ao operador obter a perspectiva de um cliente da rede, e aceder ao serviço de dados, efectuando testes periódicos de uma forma contínua e automática, em locais geograficamente relevantes para o operador, de uma forma E2E.

Conseguem-se determinar por exemplo, o tempo de estabelecimento de uma chamada de dados, se existem ou não interrupções de um determinado serviço de dados, os *throughputs* de *downlink* ou *uplink* de um ficheiro ou o tempo que demora a carregar uma página de HTTP.

Com este tipo de sistemas consegue-se aferir se a qualidade da ligação de dados e, por conseguinte, a experiência de utilização do cliente se encontra dentro de determinados parâmetros, como por exemplo se o tempo de carregamento de uma página HTTP se encontra dentro de um intervalo de tempo expectável, dependendo do tamanho da página e do *throughput* de pico definido como aceitável para esse local e perfil do utilizador. No entanto, em caso de problemas na ligação ou no acesso ao serviço, este tipo de soluções não permite identificar com precisão onde o problema se está a verificar. Se é no acesso ou na própria transmissão, ou em algum nó de *Packet Core*, ou num qualquer elemento de rede da rede de *backbone*. É aqui que a solução descrita nesta dissertação se diferencia, disponibilizando esta identificação da localização exacta do problema. Permite assim ao operador optimizar tempo na resolução do mesmo, pois consegue colocar o seu departamento de operações a resolver o problema sem perder tempo precioso a tentar identificar onde o problema poderá estar a ocorrer.

- ***L3 Performance Monitoring***

São soluções que são mais utilizadas na parte de acesso, entre o RNC e o nodeB, ou entre a SGW e o eNodeB, a que normalmente se dá o nome de *Mobile Backhaul* (MBH). Permite aferir parâmetros desse troço de rede, entre os quais, a perda de pacotes, *round-trip time*, *jitter* ou latência (1-*way*) em cada sentido.

Este tipo de solução é mais utilizado para garantir acordos de níveis de serviço, do inglês *Service Level Agreements* (*SLA)*, quando por exemplo se utilizam circuitos alugados.

No diagrama seguinte mostram-se onde as várias soluções descritas atrás se posicionam numa rede móvel, e qual o proveniência das várias origens de dados das diferentes soluções.

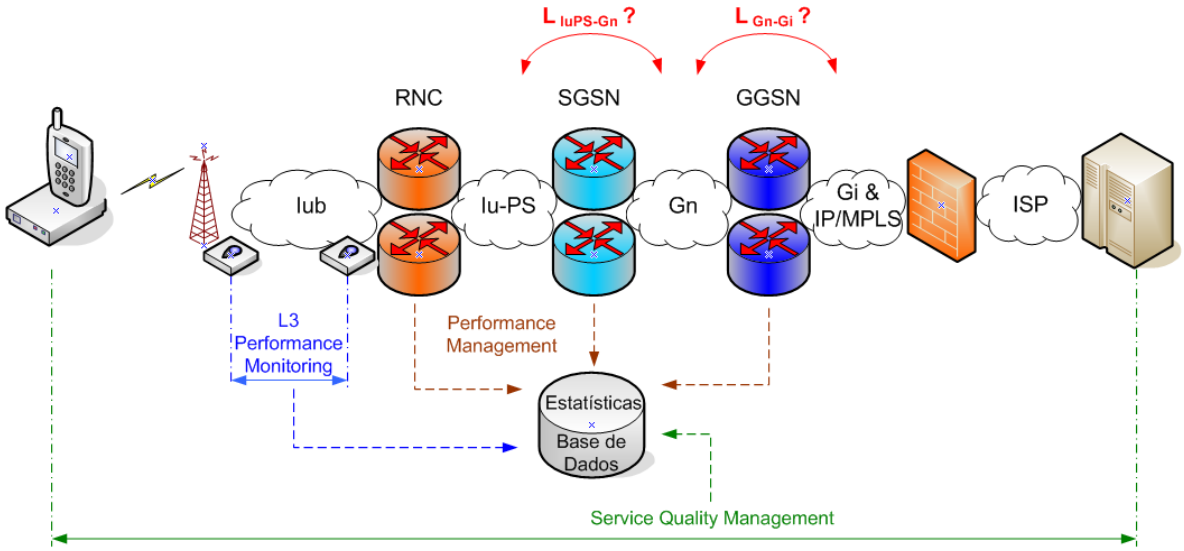

Figura 26 – Posicionamento das várias soluções numa rede móvel 3G

Estas soluções em conjunto permitem uma visibilidade do estado da rede móvel do operador considerável, mas nenhum deles, devido à singularidade dos protocolos utilizados em redes móveis 3GPP, consegue fornecer os valores estatísticos sobre a latência, que cada um dos elementos ou segmentos de rede contribuem para a latência global sentida pelo utilizador. As soluções de "*Service Quality Management*" habitualmente recolhem valores aproximados RTT de uma perspectiva E2E, recorrendo a PINGs. Estes valores têm as desvantagens já identificadas explicadas no capítulo 4.6, e têm também diluído qualquer irregularidade que possa existir nas componentes de *core* da rede móvel.

A componente de rádio sempre teve, historicamente, um peso considerável no valor total da latência, em cada um dos sentidos. Com o aparecimento do HSPA nas redes 3G/WCDMA e mais recentemente com o 4G/LTE, esta contribuição tem vindo a ser reduzida, como se pode verificar na figura seguinte [27].

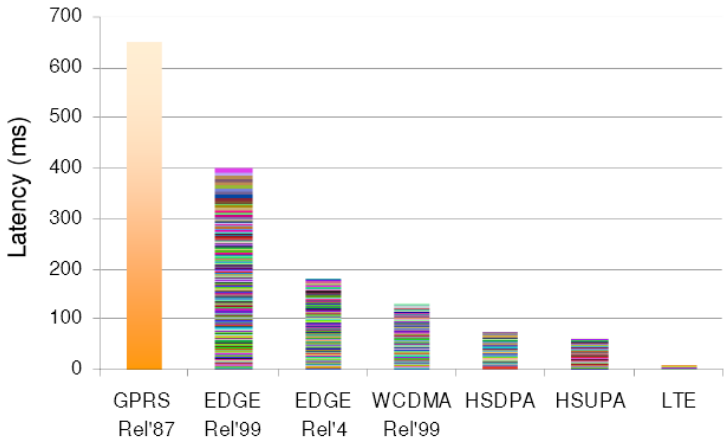

Figura 27 – Evolução dos valores Latência por tecnologia de acesso

Nas redes de 3G/WCDMA/HSPA e 4G/LTE, os valores de latência na parte de acesso rádio situam-se na ordem dos 20 a 5 ms, que comparativamente a latências verificadas na rede de *Core*, entre os 6 e o 2 ms, já não são assim tão distintas.

Desta forma é cada vez mais urgente a um operador aperceber-se, de preferência de uma forma contínua, da latência sentida em toda a sua rede, não só de uma forma E2E, mas também de uma forma segmentada, para estar apto a reagir com antecedência no caso de algo não estar de acordo com o esperado.

Nesta dissertação pretendem-se mostrar casos em que isto se torna evidente, e a mais valia para o operador em ter uma solução como a proposta, a monitorizar permanentemente a sua rede.

4.7.1 Precisão das Medições

Como foi visível na Figura 27, as latências observadas numa rede móvel são cada vez mais reduzidas. Isto torna necessário que o equipamento utilizado nas medidas tenha que ser cada vez mais preciso, logo mais dispendioso.

A precisão do relógio interno do sistema de monitorização utilizado na solução descrita nesta dissertação foi-nos garantida pelo fornecedor que tinha um *drift* máximo de 1µs por segundo. Infelizmente não nos foi possível obter os valores de referência do equipamento, tanto de o valor de PPM como da frequência de relógio.

Utilizando portanto como referência o desvio indicado, verifica-se que as medidas obtidas irão estar dentro dos valores de erro ambicionados, pois com valores no *core* a rondar os 1-2 ms,

iremos ter um erro máximo inferior a 0.1%. Como iremos ver no capítulo 7, chegamos a ter valores em determinados segmentos da rede de *core* em que as latências rondam os 150 µs. Mesmo com estes valores de latência, o erro associado à medida anda à volta dos 1%.

5. Solução Proposta

5.1 Origem da Ideia

A ideia original de tentar criar uma solução que permitisse aferir latências com maior granularidade numa rede móvel de banda larga, surgiu ao analisar o incómodo que um operador móvel multinacional estava a ter quando comparava os valores de *throughput* obtidos na sua rede, por sondas espalhadas pelo país que simulavam o cliente final, com os valores dos seus concorrentes, o qual demonstrou que seria interessante complementar estes valores com outros parâmetros que permitissem lançar alguma luz sobre a qualidade da sua rede, no qual a latência era referido como o mais interessante de se medir, pois não existia uma forma fidedigna de a obter. Com as sondas referidas atrás, sabia-se que forneciam valores interessantes da perspectiva do cliente final, mas que se porventura algo estivesse errado com o serviço, não permitiam identificar a causa do referido problema. O interesse deste operador também residia no facto de a latência ser um dos KPI's que eram analisados tanto pelos *benchmark's* nacionais como de grupo.

Observando este interesse e verificando a existência de uma lacuna na forma de obter valores sobre este parâmetro, e isto num operador europeu como mais de 20 milhões de subscritores em que existem recursos para explorar formas de optimizar a rede, surgiu a necessidade de começar a analisar as soluções já existentes no mercado que pudessem ser úteis para extrair este valor. Observou-se rapidamente que essas soluções conseguiam fornecer valores, mas essencialmente RTT e de uma forma E2E, entre um servidor e UE. Encontraram-se também soluções que incidiam em medições activas entre troços de rede mas as quais requeriam que existisse *routing* IP entre os pontos de interesse. A informação sobre as soluções que o mercado oferece já foram descritas no capítulo 2.

Pelas razões indicadas atrás e no capítulo 2, e sabendo que num operador móvel não é habitual recolherem-se valores exactos sobre o impacto que os elementos e os segmentos de rede de *core* introduziam na latência, pensou-se numa solução que pudesse satisfazer o desejo do referido operador, medir a latência de rede de uma forma E2E em ambos os sentidos, *downlink* e *uplink*, mas acrescentando algo que não é de todo habitual nas soluções existentes no mercado, a latência por troço de rede. Desta forma tentou-se acrescentar valor aos dados que podiam ser disponibilizados pela solução, pois desta forma consegue-se identificar segmentos de rede onde

um possível problema esteja a acontecer e actuar antes de os clientes sentirem qualquer tipo de degradação.

O sistema de monitorização utilizado na recolha de tráfego é um sistema que existe em qualquer operador, habitualmente para recolher tráfego de *Control-Plane* para posterior análise, mas que na nossa solução levamos um pouco mais além, incumbindo-o da tarefa de captura de um fluxo de tráfego específico de *User-Plane* em vários pontos de rede. O número de pontos de rede onde se pretende capturar o tráfego, isto é, o número de TAP's (ver capítulo 5.4), é efectivamente a maior fatia do investimento em CAPEX que tem que ser efectuado pelo operador; tudo o resto é pós-processamento que pode perfeitamente ser automatizado, minimizando os custos de OPEX para operar e gerir a solução. O facto de se utilizarem TAP's e não equipamentos externos para gerar e recolher o tráfego, está em linha com o conceito de redução de custos operacionais, pois sendo os primeiros elementos passivos, têm uma probabilidade de avaria muito mais reduzida que os segundos, logo com contratos de suporte muito mais em conta. Levou-se em consideração também o facto de se querer utilizar tráfego gerado por um utilizador final, podendo ser ou não uma sonda remota, para poder acrescentar medições de experiência de serviço percepcionada pelo cliente. Estas sondas são também ferramentas que vulgarmente já se encontram espalhadas por operadores, portanto utilizando-as como elementos geradores de tráfego, simulando o perfil de tráfego de um cliente tipo, faz todo o sentido.

Para conseguirmos fornecer medições precisas de latência com a solução que se pretendia desenhar, que era fulcral para o sucesso da solução final, tivemos que procurar no mercado equipamentos de monitorização de rede que se diferenciassem das restantes, principalmente na componente de sincronismo. A investigação levou-nos a uma *startup* Norte Americana, que disponibilizava o que pretendíamos e que se diferenciava por ter uma característica que garantia precisão às medidas efectuadas. Esta característica diferenciadora era que esta ferramenta colocava a informação sobre o instante de tempo em que pacote de *Ethernet* era capturado no próprio pacote, logo não ficávamos dependentes nem da distância do local de captura ao servidor onde o tráfego seria armazenado, nem da qualidade de relógio desse mesmo servidor, que como se desconfia, não são reconhecidos por utilizarem relógios precisos pois não foram construídos com esse intuito.

5.2 Descrição de Alguns Componentes da Solução

A solução criada utiliza como origem do tráfego sessões de dados iniciadas por utilizadores com SIM's de teste do operador, colocados em *modems*, como *dongles* ou dispositivos *handheld*, ou então por sondas, como as existentes em sistemas de *Service Quality Management*. Tanto os *modems* como as TAP's deverão ser colocados em pontos considerados estratégicos na rede, para medirem áreas de cobertura em segmentos da rede que sejam considerados relevantes. Exemplos de áreas relevantes são zonas empresariais, comerciais ou os centros das cidades. Pode ser também interessante aferir segmentos de rede na parte de acesso, que sejam circuitos alugados, os quais sejam relevantes controlar o SLA (do Inglês, "*Service Level Agreement*") acordado ou então a transmissão entre *sites* primários do operador, onde se situam elementos de rede como os SGSN's, GGSN's ou então o POP's do ISP.

Capturando todo o fluxo de dados criado por aplicações específicas, como HTTP, FTP ou então efectuando sessões de tráfego UDP com uma elevada cadência e aleatoriedade, em diferentes segmentos de rede de uma forma síncrona, consegue-se correlacionar *à posteriori* a informação recolhida e disponibilizar uma gama de indicadores relevantes para aferir a qualidade com que um determinado serviço é disponibilizado pela rede ao utilizador final, como também aferir a própria qualidade de um determinado pedaço da rede, ou então podendo chegar mesmo a se poder aferir a própria qualidade de um elemento de rede.

Os pacotes de IP são recolhidos por um sistema de monitorização que coloca em cada um dos pacotes capturados um selo com o instante em que o pacote foi capturado. Os pacotes são recolhidos se correspondem ao filtro colocado em cada uma das interfaces dos quais interessa capturar o tráfego. Com a informação do selo com o instante de tempo da captura, permite que se possam fazer correlações como as seguintes:

Latência: Medição do tempo que um pacote demora a fluir de um ponto A a um ponto B.

Packet Loss: Permite saber se um determinado pacote que foi detectado num ponto A, ainda é observado num ponto B.

Jitter: Permite saber o intervalo de tempo entre cada pacote de determinado fluxo de pacotes enviado de uma fonte precisa.

Como já clarificamos no início desta dissertação, iremos focar-nos somente no primeiro indicador, a latência.

A solução criada foi inicialmente pensada e testada em redes 3G/WCDMA/HSPA, em que todas as interfaces são IP sobre *Ethernet*. No entanto, os mesmos princípios aplicam-se à arquitectura das redes 4G/LTE, desde que não se utilize IP-Sec no interface S1-U, pois neste caso, estando este interface encriptado, não seriam possível correlacionar os pacotes de IP capturados nessa interface com os demais, capturados em outras interfaces.

É de salientar que a solução utiliza um sistema de monitorização com uma característica diferenciadora. Aquando da captura do tráfego, esta inclui em todos os pacotes de *Ethernet* que correspondem ao filtro aplicado nesse interface, um campo com a informação sobre o instante de tempo em que foi capturado. Esta característica irá ser explicada no capítulo seguinte.

A solução é constituída por três grandes blocos, que resumidamente se encontram definidos na Figura 28 a cinzento. Nos capítulos seguintes iremos detalhar cada um destes blocos.

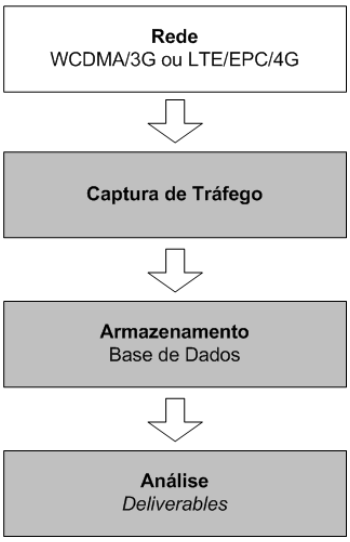

Figura 28 – Definição do processo

5.3 Arquitectura da Solução

Ao se arquitectar esta solução pensou-se de raiz em se construir uma solução que fosse versátil e que pudesse ser utilizada em qualquer rede, desde as redes 3G/WCDMA/HSPA como também em redes 4G/LTE/EPC. No seguinte diagrama é apresentada uma perspectiva de alto nível de uma possível arquitectura desta solução, inserida numa rede 3GPP 3G/WCDMA/HSPA.

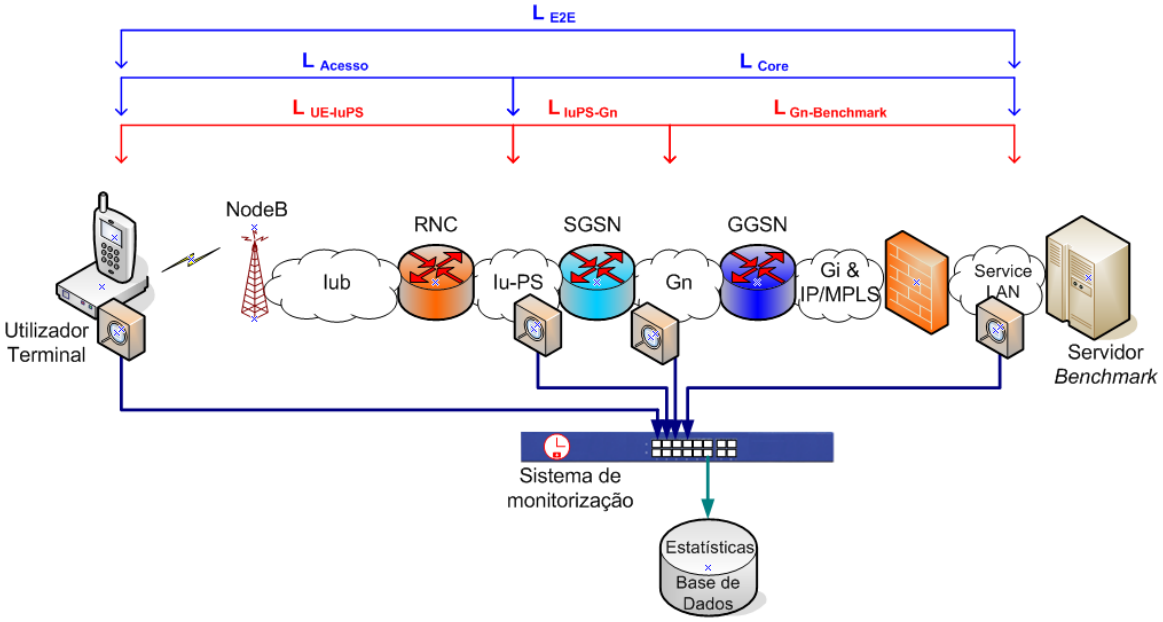

Figura 29 – Perspectiva de alto nível da solução, WCDMA/3G

Neste diagrama são visíveis as diversas redes e elementos de rede que se encontram definidos no 3GPP TS 23.060 [28], numa arquitectura GPRS baseada em interfaces Gn/Gp. Mostram-se também que são disponibilizados 3 segmentos de rede, de onde se podem recolher amostras e medir as latências associadas, além claro, da latência entre o UE e o Servidor utilizado como *benchmark*.

No caso de uma rede 4G/LTE/EPC, uma possível disposição das *TAP's* na rede seria a seguinte.

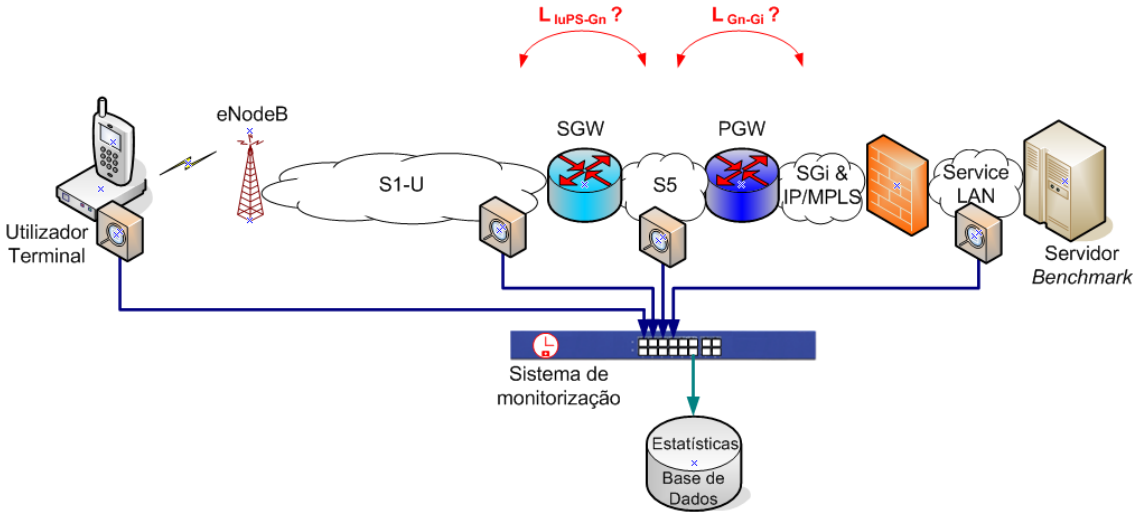

Figura 30 – Perspectiva de alto nível da solução – LTE/4G, EPC

No capítulo anterior referiu-se a granularidade que era possível alcançar com esta solução. Referiu-se também a ambição e capacidade de investimento. Isto é visível no diagrama seguinte em que se mostra como exemplo uma arquitectura mais rica em TAP's, mas também um pouco mais complexa e com maior impacto no investimento financeiro necessário para a criar. Podemos ver que, comparativamente ao diagrama indicado na figura 13, conseguimos obter o dobro da granularidade, 6 segmentos de rede monitorizados.

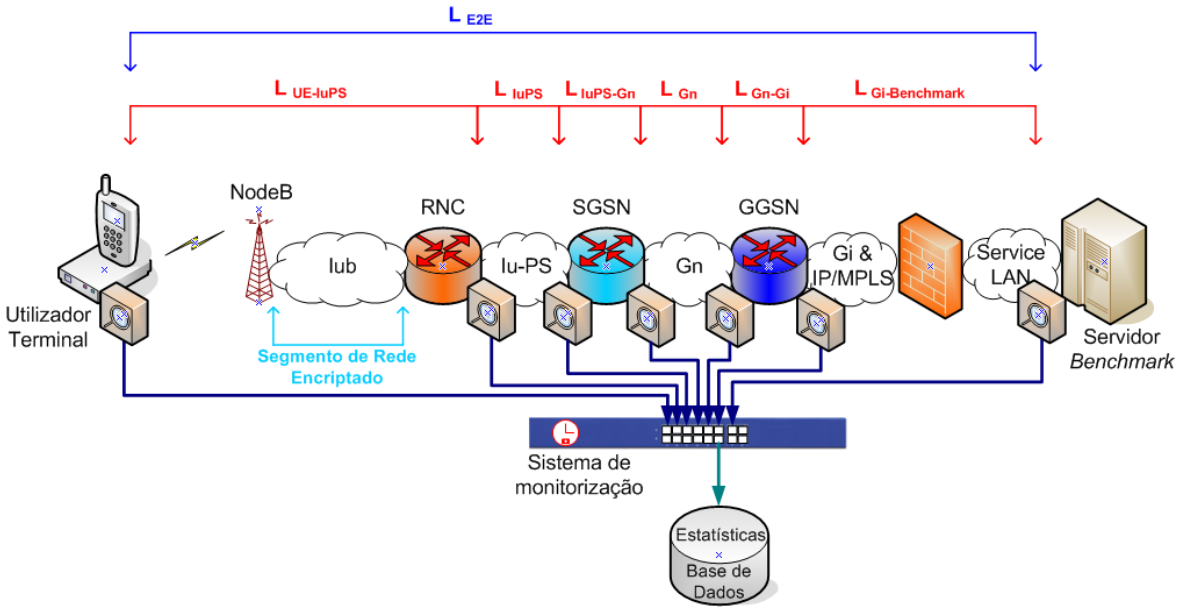

Figura 31 – Perspectiva de alto nível da solução, WCDMA/3G

O servidor *benchmark* indicado nas figuras anteriores foi colocado numa rede de serviços na cidade A, controlada pelo próprio operador, evitando desta forma a utilização da Internet. O racional por detrás desta decisão foi o de avaliar comportamentos de segmentos de rede ou de

equipamentos que sejam relevantes para o operador, em que ele possa intervir para optimizar ou melhorar o seu comportamento, em caso de necessidade.

Devido a se ter possibilidade de encaminhar o tráfego para um GGSN colocado numa cidade distante do local onde se efectuaram os testes, fizemos uso deste facto para aferir a penalidade que advém quando um operador decide balancear o tráfego dos utilizadores por GGSN's distantes de onde o tráfego é originado.

Um diagrama simplificado da rede é mostrado na seguinte figura, indicando o fluxo de dados entre os elementos de rede situados na cidade A, a verde mais escuro, e o fluxo de dados que é entregue na cidade B e regressa à cidade A, a verde mais claro.

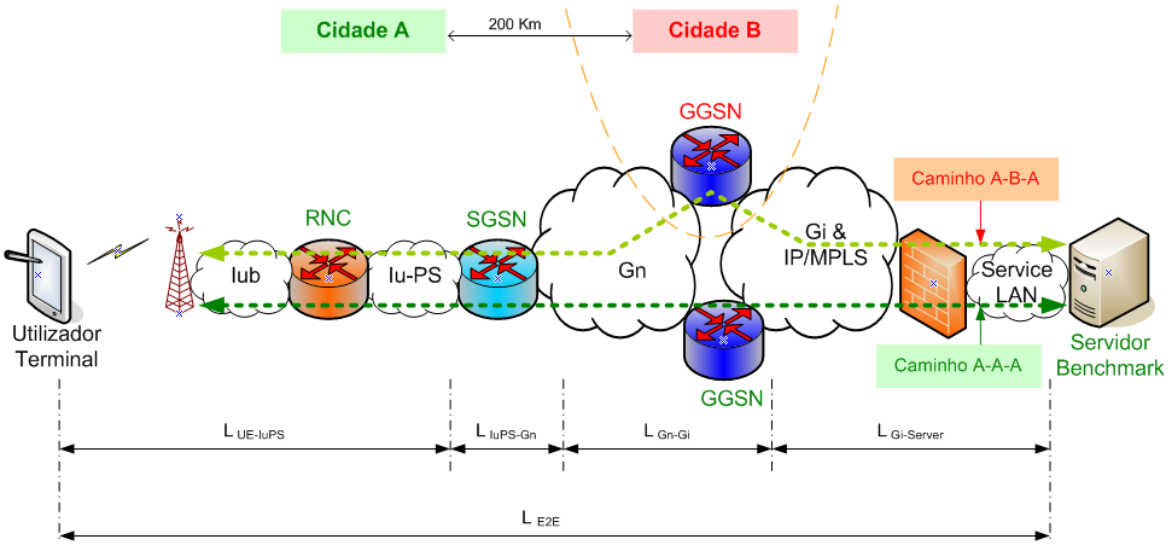

Figura 32 – 3GPP *User Plane* - UTRAN com Gn - Diagrama Simplificado com os Fluxos

Com a informação recolhida nos diversos pontos de captura e depois de tratada, é possível fornecer ao operador uma fotografia sobre o comportamento global da sua rede, as possíveis flutuações dos diversos KPIs por segmento de rede, por exemplo aumentos esporádicos de latência num determinado segmento da rede ou numa determinada hora do dia, e portanto identificar possíveis impactos na qualidade de experiência do utilizador. Como esta informação pode ser recolhida de uma forma permanente, o operador pode armazenar um histórico com o qual permite observar como esses KPI's evoluem com as alterações diárias efectuadas na rede tanto na parte de acesso ou de *Core*, quando se altera a arquitectura da mesma, quando se introduz uma nova versão de *software* num determinado componente de rede ou mesmo quando de introduz ou se substitui um elemento de rede por um novo de um outro fornecedor.

5.4 Captura de Tráfego

Nesta fase definem-se quais as interfaces que serão relevantes capturar e estruturam-se os segmentos de rede a analisar. Aqui o limite é a granularidade desejada que melhor vá ao encontro dos seus desejos e ambições.

Para se recolher o tráfego de um determinado interface recorrem-se a TAP's. Estas são equipamentos passivos de rede que replicam num porto, denominado de porto de monitorização, o tráfego que fluí entre dois elementos de rede, como por exemplo entre o GGSN e o *switch* de *layer* 2. Na figura seguinte exemplifica-se em que parte da rede se posiciona a TAP.

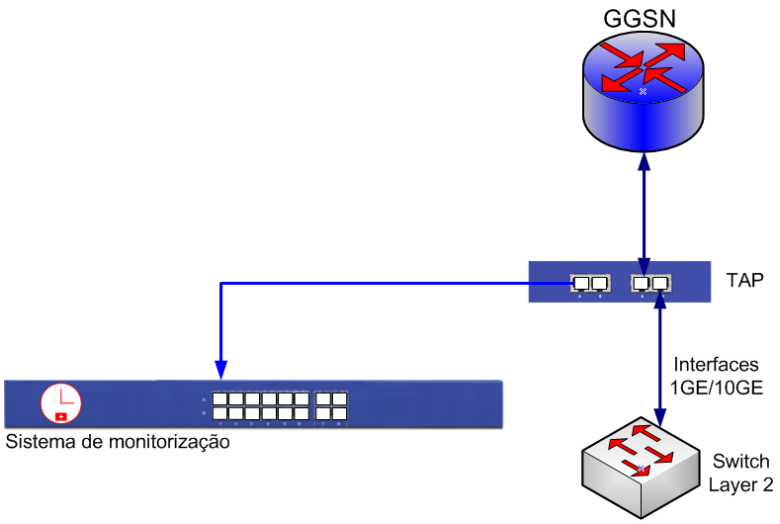

Figura 33 – TAP, localização na Rede

Os datagramas de *Ethernet* capturados pelo sistema de monitorização, que correspondem às condições definidas nos filtros aplicados a uma determinada interface, são enriquecidos com informação complementar. Esta informação consiste, como já indicado anteriormente, no instante em que *o* datagrama foi capturado e que consiste no dia, na hora, no minuto, no segundo e por fim informação em nano-segundos recolhido do relógio interno do sistema de monitorização, que tem uma precisão muito maior que um servidor tradicional. Existe também informação sobre qual o porto do sistema de monitorização de onde o datagrama foi recolhido, para desta forma saber-se de que interface o tráfego foi recolhido. Sem esta informação não seria possível fazer a correlação da informação. Na Figura seguinte temos um esquema do pacote original de *Ethernet* e de como esse mesmo pacote fica depois de ter sido capturado pelo sistema de monitorização.

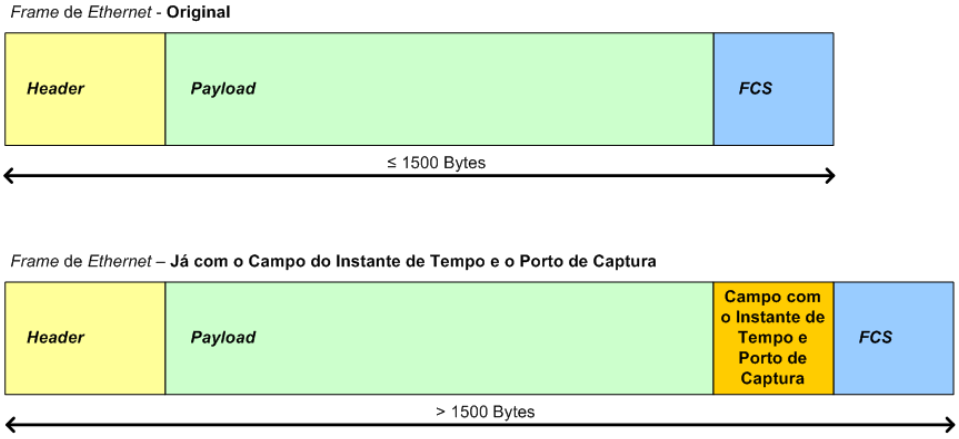

Figura 34 – Localização do Campo adicionado pelo Sistema de monitorização no *Frame Ethernet*

Todo o fluxo de tráfego capturado das diversas interfaces é encaminhado para um servidor que irá ser utilizado para armazenar a informação recolhida. De notar que, para dimensionamento desta interface, terá que se entrar em consideração com a velocidade máxima alcançada pelo subscritor final que depois terá que ser multiplicada pelo número de interfaces de onde o tráfego irá ser recolhido.

Por exemplo, se a velocidade máxima estimada para a sessão for de 50 Mbit/seg, devido a boas condições de rádio e ao perfil correcto definido no HLR, e com captura em 4 pontos específicos de rede, o valor enviado para o servidor será de, 50 x 4 = 200 Mbit/seg. Como nota, o valor recomendado para a ocupação de um porto 1000BaseT é 80%, logo 800 Mbit/seg. No entanto, pode haver necessidade devido às velocidades alcançadas de haver mais do que um porto de 1Gbps para acomodar todo o tráfego capturado no servidor.

5.5 Armazenamento dos Dados

O servidor tem nesta solução um papel importante pois é ele que armazena os diversos fluxos de dados recolhidos da rede, e para não se incorrer em aferições erradas de perda de pacotes, tem que ter capacidade suficiente para escrever no disco rígido os dados que lhe são enviados pelo sistema de monitorização.

Aqui é indispensável ter instalado o *libpcap*, que é uma livraria utilizada para capturar tráfego de redes IP. É necessário também ter a correr, na altura das capturas, uma aplicação de análise de pacotes como o *tcpdump* ou *Wireshark*. Estas aplicações são essenciais para salvaguardar toda a informação recolhida da rede pelo sistema de monitorização.

Como o sistema de monitorização adiciona ao datagrama de *Ethernet* o campo extra com a informação do tempo e do porto, ultrapassando o seu valor tradicional, é também necessário que este servidor tenha interfaces de rede, em inglês *Network Interface Card* (NIC), que suportem *Jumbo Frames*. Desta forma o servidor já consegue interpretar os pacotes e acomodá-los no seu disco rígido. Sem esta característica nas interfaces de rede, todos os pacotes com tamanho superior a 1500 *bytes* seriam simplesmente descartados pela mesma.

Por questões de economia de equipamento, este servidor pode também ser utilizado para correr o servidor de MySQL. Sobre o último assunto iremos falar no capítulo seguinte.

5.6 Base de Dados

Toda a análise irá ser efectuada sobre uma base de dados MySQL. Escolheu-se este tipo de base de dados pois é *Open Source* o que faz com que não tenhamos despesas com licenciamento, mantendo então os custos da solução controlados.

Armazenando os dados numa base de dados acarreta consigo as vantagens descritas a seguir.

- Simplificar a informação recolhida.

Os dados recolhidos são massivos, tanto em volume de pacotes como em informação detalhada sobre cada um destes pacotes. Filtrar a informação relevante torna-se essencial, tanto para simplificar e para optimizar as tabelas, como para tornar mais célere o acesso à informação e não sobrecarregar a base de dados com informação irrelevante.

- Ter um histórico

Pode-se acompanhar a evolução das optimizações na rede, como da introdução de novas funcionalidades, de novas revisões de *software* nos elementos de rede, ou até mesmo comportamentos com diferentes fornecedores.

- Disponibilizar *Benchmark's*

Permite fazer um *benchmark* entre diferentes operadores, e até mesmo diferentes fornecedores ou cenários. Isto depende obviamente da qualidade e da quantidade de dados que são possíveis popular na base de dados.

Com esta informação pode-se enriquecer o que se pode entregar com estes serviços, pois permite posicionar a análise feita a um operador, num universo de operadores anónimos mas semelhantes.

5.7 Análise dos Dados

Para se poder efectuar a análise dos dados existentes na base de dados de uma forma eficiente produziram-se *queries* em MySQL, tanto para a destilar a informação original como para apresentar os dados já processados, isto é, os diversos valores de latência, obtidos nos diversos cenários analisados.

Na fase de destilação da informação foram desenvolvidos *Store Procedures* em MySQL para extrair o essencial dos dados. Desta forma suavizou-se a informação recolhida, colocando em tabelas específicas por KPI os dados estruturados. Desta forma organizou-se a informação de forma a se poder ter, por exemplo, a latência por amostra, a latência por segundo, a função distribuição de probabilidade, vulgo PDF ou em inglês *Probability Distribution Function*, e por fim a função distribuição cumulativa de probabilidade, CDF ou em inglês *Cumulative Distribution Function*. Esta informação é organizada por operador, cenário ou condições do teste.

Os cálculos efectuados são baseados nas diferenças temporais das capturas nas diversas interfaces do mesmo pacote, utilizando para tal a referência de relógio incluída no pacote *ethernet* pelo sistema de monitorização. Por exemplo, para o cálculo da latência entre o *benchmark server* e o UE em *downlink*, era obtido o valor do tempo de um determinado pacote na interface do UE, que era a última interface que o pacote percorria, subtraindo o valor temporal na interface do *benchmark server* desse mesmo pacote, sendo este a primeira interface que o pacote fluía. Para obter o valor de latência em *uplink*, o mesmo processo, mas de forma inversa era utilizado. De seguida apresentamos a fórmula genérica utilizada nos dois sentidos do tráfego.

$$Latência_{DL(A-B)} = \Delta(Inst_{CapturaB} - Inst_{CapturaA})$$

Fórmula 2 – Cálculo do valor da Latência de *Downlink* da Interface A à Interface B

$$Latência_{UL(B-A)} = \Delta(Inst_{CapturaA} - Inst_{CapturaB})$$

Fórmula 3 – Cálculo do valor da Latência de *Uplink* da Interface B à Interface A

Para a apresentação dos dados decidiu-se utilizar como ferramenta de *Reporting* a aplicação *Business Objects*. Esta aplicação não é *freeware* mas, como já disponhamos de uma licença e como já tínhamos à-vontade com a mesma, decidiu-se utilizá-la para efectuar toda a parte gráfica

e de apresentação. No entanto, existem soluções *Open Source* que permitiriam efectuar um trabalho com a mesma qualidade, as quais também têm interfaces para bases de dados MySQL.

5.8 Posicionamento da Solução em termos das ofertas de Mercado

Esta solução posiciona-se na gama dos "*Service Assurance Systems*", cobrindo na plenitude dois dos seus subdomínios, o "*Performance Monitoring*", mas também podendo substituir alguns dispositivos do subdomínio das "*Probe Systems*".

O valor global do mercado previsto pelos analistas da *Analysys Mason* para 2015 nos "*Service Assurance Systems*" [29] foi de 3.5 BUSD. Para o "*Performance Monitoring*" e "*Probe Systems*" os valores são de 643 MUSD e 1.330 MUSD, respectivamente. O CAGR entre 2010 e 2015 foi estimado que irá rondar os 9.4%, pelo mesmo relatório. De referir também que a grande fatia do investimento, cerca 53%, irá ser feita por operadores móveis, principalmente devido à implementação de redes 4G/LTE, que se estima que chegue a gerar receitas de 1.9 BUSD em 2015. Isto indica portanto que é uma área em que existe interesse no mercado, e que tem uma margem de crescimento relevante de onde se podem retirar significativos benefícios financeiros.

A grande maioria dos operadores móveis tem à sua disposição sistemas de monitorização que são utilizados para recolher informação relativa à sinalização gerada pelos seus clientes. Esta informação é armazenada, permitindo ao operador explorar o que se passou com essa sinalização e identificar a causa de por exemplo, quando um cliente se queixa de não conseguir aceder a um determinado serviço. Estas soluções cobrem habitualmente diversas redes, como por exemplo na parte de acesso o Iu-PS, entre o RNC e o SGSN, e na parte de *Core* a Gn, entre o SGSN e o GGSN/PGW. A parte de *User-Plane*, isto é o tráfego real dos subscritores, não é de todo monitorizada ou recolhida. Isto deve-se logicamente a constrangimentos de capacidade de processamento e a espaço de armazenamento, mas também devido à forma como o licenciamento destes produtos é efectuada e cobrada aos operadores, que está indexada à quantidade de tráfego monitorizado.

Outra ferramenta que é vulgarmente utilizada pelos operadores móveis são as sondas, que geram tráfego em instantes de tempo específicos, e que tentam aferir a disponibilidade e qualidade de vários serviços disponibilizados pelo operador.

A solução que propomos tem como principal objectivo ultrapassar algumas limitações existentes nos sistemas de monitorização actuais, podendo mesmo em grande parte da rede substituir na totalidade algumas dessas soluções. Como tem o intuito de utilizar tráfego gerado

por sondas colocadas no lugar de clientes finais, tem a capacidade de complementar estas soluções e absorver os seus dados da rede de uma forma síncrona e granular, estando somente dependente do interesse e disponibilidade de CAPEX no investimento em *TAP's* distribuídas pela rede. Tem também que se entrar em conta com a diversidade de portos disponíveis no sistema de monitorização.

No capítulo seguinte vamos mostrar vários diagramas com possíveis arquitecturas e granularidades.

6. Cenários de Teste

6.1 Qualidade da Rede

Para se aferir a qualidade de uma rede interessa que se consigam medir os diversos KPI's e que os mesmos se encontrem dentro de limiares predefinidos. Nos documentos dos 3GPP, 3GPP TS 23.107 [17] e 3GPP TS 23.203 [30], encontram-se definidos respectivamente, os valores que cada *bearer* deve cumprir, tanto para 3G como para 4G.

Os valores definidos para os diversos tipos de classes de tráfego 3G encontram-se definidos na seguinte tabela, na qual realçamos nas células azuis os valores de latência máxima que a rede tem que cumprir.

De referir que os 3GPP definem este valor máximo de latência como o percentil 95 da distribuição de todos os pacotes transmitidos durante o tempo de vida do *bearer*.

Tabela 8 – Intervalo de valores definidos para os diversos *bearer* 3G

Traffic class	**Conversational class**	**Streaming class**	**Interactive class**	**Background class**
Maximum bitrate (kbps)	<= 256 000	<= 256 000	<= 256 000	<= 256 000
Delivery order	Yes/No	Yes/No	Yes/No	Yes/No
Maximum SDU size (octets)	<=1 500 or 1 502	<=1 500 or 1 502	<=1 500 or 1 502	<=1 500 or 1 502
SDU format information				
Delivery of erroneous SDUs	Yes/No/-	Yes/No/- (6)	Yes/No/-	Yes/No/-
Residual BER	$5*10^{-2}$, 10^{-2}, $5*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}	$5*10^{-2}$, 10^{-2}, $5*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}	$4*10^{-3}$, 10^{-5}, $6*10^{-8}$ (7)	$4*10^{-3}$, 10^{-5}, $6*10^{-8}$
SDU error ratio	10^{-2}, $7*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}	10^{-1}, 10^{-2}, $7*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}	10^{-3}, 10^{-4}, 10^{-6}	10^{-3}, 10^{-4}, 10^{-6}
Transfer delay (ms)	**100 – maximum value**	**300 – maximum value**		
Guaranteed bit rate (kbps)	<= 256 000	<= 256 000		
Traffic handling priority			1,2,3	
Allocation/Retention priority	1,2,3	1,2,3	1,2,3	1,2,3
Source statistic descriptor	Speech/unknown	Speech/unknown		
Signalling Indication			Yes/N	

Os valores definidos para os *bearers* 4G encontram-se definidos a seguir. As células azuis continuam a realçar os valores de latência máxima que a rede tem que cumprir para os diferentes QCIs.

Tabela 9 – Características de Transmissão dos diversos *bearers* de 4G/LTE

QCI	Tipo de Recurso	Prioridade	Agregado da Latência do Pacote	Taxa de Perda de Pacotes	Exemplos de Serviços
1	GBR	2	100 ms	10^{-2}	*Conversational Voice*
2		4	150 ms	10^{-3}	*Conversational Video* (*Live Streaming*)
3		3	50 ms	10^{-3}	*Real Time Gaming*
4		5	300 ms	10^{-6}	*Non-Conversational Video* (*Buffered Streaming*)
5	Non-GBR	1	100 ms	10^{-6}	IMS *Signalling*
6		6	300 ms	10^{-6}	Video (*Buffered Streaming*) TCP-*based* (e.g., www, e-mail, *chat*, ftp, p2p file *sharing*, *progressive video*, etc.)
7		7	100 ms	10^{-3}	*Voice*, Video (*Live Streaming*) *Interactive Gaming*
8		8	300 ms	10^{-6}	Video (*Buffered Streaming*) TCP-*based* (e.g., www, e-mail, *chat*, ftp, p2p file *sharing*, *progressive video*, etc.)
9		9			

Estes serão os valores que iremos ter em consideração quando analisarmos os valores obtidos das redes analisadas. De referir que todos os *bearers* utilizados nos testes foram da classe de tráfego *Interactive*, pois é a classe utilizada pelos subscritores normais de rede. À data destes testes não era de todo normal encontrarem-se operadores que disponibilizassem serviços que utilizassem as classes *Streaming* on *Conversational*. De qualquer forma, os valores obtidos nos testes irão ser confrontados com os valores definidos nos 3GPP, pois na classe *Interactive* não é definido qualquer valor de latência ou como é referido no 3GPP TS 23.107, *Transfer Delay*.

6.2 Arquitectura do Cenário de Testes

O cenário montado foi idêntico à Figura 35, no qual extraímos o tráfego em quatro pontos da rede. No UE, ou *End-User*, no Iu-PS, ou IuU, na Gn e na rede de serviços, já a montante da Gi, onde se encontrava o servidor de testes também denominado por *Webserver*. De notar que na interface que se situa entre o UE e o Iu-PS, consiste em dois meios de transmissão, um físico que se localiza entre o RNC e o NodeB e do qual fazem parte uma infra-estrutura constituída por fibra ou micro-ondas, *routers* e *switches*. A outra parte consiste no interface ar, entre o NodeB e o UE.

Recorreram-se a testes activos iniciados por um *Laptop* no qual se correram diversas aplicações, como um cliente de FTP e uma aplicação que gerava pacotes de UDP. O *Laptop* encontrava-se conectado a um *router* 3G que por sua vez interagia com a rede de banda larga do operador. O diagrama é mostrado na figura seguinte.

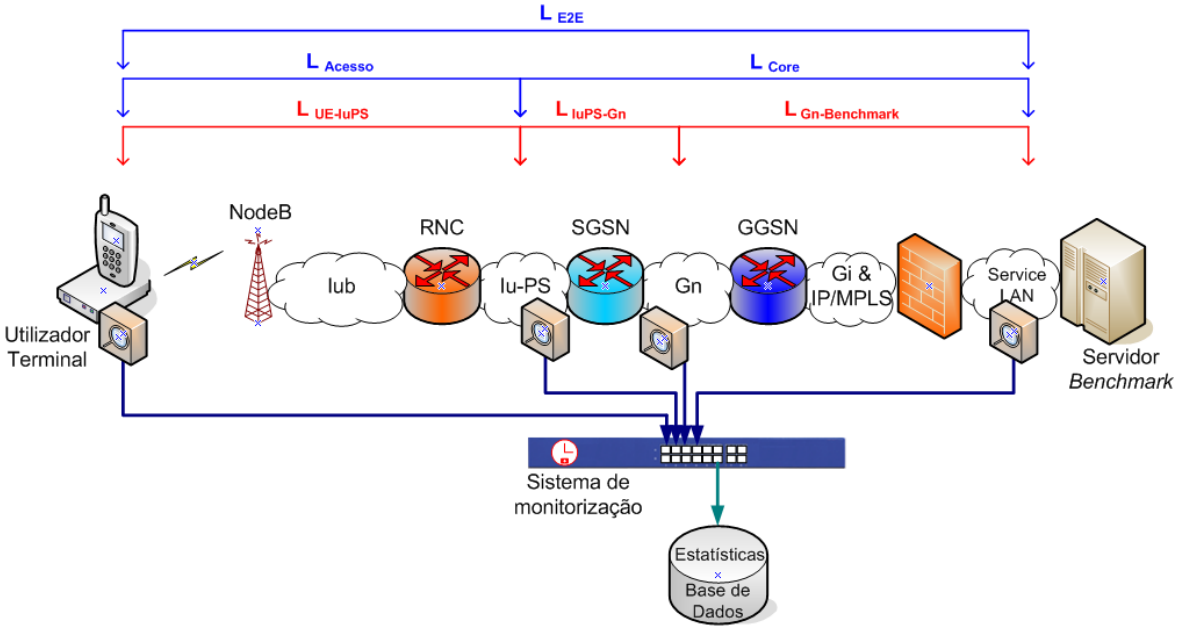

Figura 35 – Perspectiva de alto nível da solução, WCDMA/3G

6.3 Testes

Para se medir a qualidade das redes analisadas optou-se por se efectuarem diversos testes com o intuito verificar a rede do operador em diferentes condições e com a intenção de recolher dados que fossem úteis na análise global da própria rede.

Efectuaram-se testes estáticos, pois a componente com maior interesse eram os valores observados nos segmentos de *core* pois era aí que faltava visibilidade, e que portanto, testes de mobilidade não iriam adicionar valor aos resultados obtidos. O tamanho dos pacotes foi fixado a um limite máximo de 1460, tanto para os testes de TCP como de UDP, para evitar fragmentação nos segmentos de rede que sofriam de encapsulamento de GTP. O tamanho máximo definido nas redes móveis é de 1500 bytes, como definido em [28]. Este valor tem em consideração somente as camadas de *Transmission Control Protocol*/*Internet Protocol* (TCP/IP) ou *User Datagram Protocol*/*Internet Protocol* (UDP/IP) e a de *payload* do utilizador. Como além destas camadas, ainda são necessárias acomodar os cabeçalhos de GPRS *Tunnel Protocol* (GTP) que tem 8 bytes, e os de UDP com 8 bytes e de IP com 20 bytes, ambos pertencentes à camada de transporte. Decidiu-se utilizar o valor de 1460 bytes no tamanho máximo dos pacotes utilizados nos testes de forma a evitar fragmentação dos pacotes nas interfaces com encapsulamento GTP.

Os testes de UDP foram fixados a um ritmo de 3 pacotes por segundo, de forma se conseguir uma granularidade abaixo do segundo, mas que não criasse demasiados dados nos ficheiros que eram utilizados no pós-processamento.

Recolheram-se as seguintes informações dos testes:

- Valores médios e mínimos, de um universo de 10 repetições;
- Valores médios de RTT de cada um dos testes para permitir posicionar a rede do operador sobre o gráfico da Figura 22 – Valores Teóricos *Throughput* vs. *Packet Loss,* com RTT fixo";
- Evolução da latência por segmento de rede, com as suas componentes do acesso rádio e nos dois troços de *core*. Permite uma representação visual da latência verificada nos troços em questão;
- Função distribuição cumulativa de probabilidade (CDF) de uma forma E2E e por segmento de rede.

Como foi referido acima, no capítulo 6.1, no 3GPP indica-se que o valor de referência das latências numa determinada classe de tráfego deverá ser calculado como o percentil 95 da

distribuição de todos os pacotes transmitidos durante o tempo de vida do *bearer*. De notar que, nas tabelas que colocamos em baixo com os valores de cada um dos testes, não indicamos os valores máximos obtidos, pois podia desviar a atenção para o essencial. No entanto não ignoramos os valores mínimos, pois estes põem a descoberto uma informação muito relevante para um operador móvel, que é o valor mínimo de latência que a sua rede consegue disponibilizar. Esta informação é muito importante para determinados serviços, como o *Gaming*.

De referir que ambas as redes suportavam, à altura destes testes, 21Mbps em *downlink*, mas que devido a impossibilidades em efectuar optimizações do rádio, não se conseguiram velocidades de transferências tão altas como as desejadas. No entanto ressalva-se que, para o que se tinha como objectivo inicial, este facto não teve efectivamente impacto nos resultados obtidos.

O racional por detrás de cada um dos testes irá ser descrito de seguida.

6.3.1 Testes com Tráfego Concentrado na Mesma Cidade

Com estes testes tentou-se aferir qual o comportamento da rede quando de efectuava uma sessão de FTP para um servidor na rede de serviços do operador, efectuando um *download* ou um *upload* de um ficheiro com alguma dimensão, 30MB (*Mega Bytes*), para confirmar o comportamento da rede em termos de latência E2E e por segmento de rede.

Da sessão de FTP iremos utilizar o fluxo de TCP no porto 21 que é utilizado na extracção, quando se faz o *donwload*, ou inserção, quando se faz o *upload*, do ficheiro. Desta forma podemos aferir o comportamento da camada de TCP desta sessão em termos de latência dos dois sentidos do tráfego.

O tráfego nestes testes encontra-se concentrado na mesma cidade, portanto o SGSN e o GGSN estão co-localizados na mesma central do operador. O interesse aqui era o de não incluir uma nova variável na equação, que seria a distância entre os nós. No teste explicitado no capítulo 6.3.2 abaixo colocou-se o tráfego a apontar para um GGSN distante, e aqui sim quisemos sentir esta variável.

Este teste e o referido no capítulo seguinte permitem colocar a sessão do utilizador sobre a Figura 22 que nos permite determinar a velocidade máxima que este subscritor teoricamente conseguirá obter no local onde está a navegar. Desta forma obtém-se uma medida indirecta mas bastante precisa da qualidade do serviço, e logo da experiência de utilização que esse subscritor está a ter. Esta é a primeira vantagem desta metodologia.

6.3.2 Testes com Tráfego Ancorado num GGSN Situado numa Cidade Distante

Aqui continuamos a utilizar o mesmo princípio e moldes utilizados no capítulo 6.3.1, mas desta vez apontamos e ancoramos o tráfego num GGSN que ficava a cerca de 250 km de distância, da cidade onde estava localizado o SGSN. Desta forma, conseguia-se aferir a penalidade na latência final e parcelar, e ao mesmo tempo o comportamento desse mesmo segmento de rede.

6.3.3 Testes com Captura Durante 24 Horas

O principal objectivo deste teste foi o que verificar o comportamento da rede ou dos segmentos ou elementos de rede, logo nas filas de espera desses mesmos elementos, durante um dia de semana, desde a altura em que a carga da rede é baixa, isto é, quando o número de utilizadores é reduzido, mas também verificar esse mesmo comportamento quando a carga está no seu pico. Desta forma valida-se se a rede no seu todo se comporta de uma forma previsível e linear ou então, se existem algumas dependências da carga de rede em algum desses elementos ou segmentos de rede.

Se se encontrar alguma dependência da carga de rede, esse facto é um sinal que algo de errado existe com um elemento de rede num determinado segmento, o que deverá despoletar uma acção da parte do operador para resolver a situação o mais rapidamente possível. Esta pode-se considerar a segunda grande vantagem da metodologia apresentada.

De notar que este teste foi feito com tráfego de *uplink*, pois para se permitir o mesmo teste em *downlink* pequenas adaptações no GGSN iriam seriam necessárias, o que não se enquadravam com as actividades operacionais que se encontravam a decorrer em paralelo. Pelo que irá ser demonstrado nos capítulos 7.1 e 7.2, o comportamento em ambos os sentidos nos segmentos do *core* são bastante simétricos, portanto na sua essência o facto de se efectuar este teste somente em *uplink* não altera o objectivo do teste. No rádio, esta simetria já não se verifica.

7. Resultados

Nos sub-capítulos que se seguem iremos apresentar os resultados dos testes efectuados em duas redes móveis de banda larga europeias. Estes sub-capítulos têm o intuito de investigar o comportamento dos diversos segmentos de rede quando se efectuam os testes activos, explicados no capítulo anterior.

Os dados dos sub-capítulos 7.1 e 7.2, foram obtidos numa rede com mais de 5.5 milhões de subscritores em que o operador tem uma dimensão local, mas que no país em que opera é líder de mercado.

Os valores obtidos no capítulo 7.3 foram obtidos num operador multinacional que, no país no qual foram efectuados os testes, tem a segunda posição do mercado com mais de 4 milhões de subscritores.

7.1 Testes com Tráfego Concentrado na Mesma Cidade

Foi efectuada uma média e analisados os resultados estatísticos de 10 repetições, o que deu um total de 227.500 amostras analisadas.

7.1.1 Testes de *Download* – SGSN e GGSN Co-localizados

Neste teste verificou-se uma latência média em *downlink* (DL) do servidor localizado na rede de serviços, para o terminal, UE, de 197,72 ms, como indicado na Tabela 10. Na Figura 37 é visível que o percentil 95 se situa próximo dos 570 ms, fazendo com que o valor médio seja sensivelmente um terço deste. O valor médio permite cumprir os requisitos de latência para a classe de tráfego de *Streaming*, ou qualquer serviço de vídeo não conversacional. No entanto não permitia qualquer serviço de voz, com ou sem vídeo, nem mesmo qualquer tipo de g*a*ming. Olhando para o valor do percentil 95, verifica-se que nenhum serviço, à excepção do *web browsing* ou qualquer serviço de vídeo *buffering streaming* sem ser em tempo real, podem ser acomodados por este *bearer*.

O valor mínimo de latência verificado mostra que a rede consegue disponibilizar valores de 14 ms, isto é, 14 vezes inferiores ao valor médio determinado. Isto mostra que a rede tem no entanto capacidade de conseguir entregar valores muito baixos de latência, mais reduzidos dos que são necessários para *Real Time Gaming*.

De realçar que a componente de *core* tem um comportamento muito baixo de latência, o qual se pode observar na Tabela 10 e no CDF da Figura 38, responsável por 0,6% do valor global, em média. Tem também um comportamento bastante previsível, como se verifica na Figura 36. De notar que nesta mesma figura, na interface de acesso, a verde, entre o IuU e o UE, se verificam alguns efeitos de armazenamento de pacotes nas filas de espera do RNC, que são compreensíveis e espectáveis devido às adaptações que este elemento de redes tem que fazer às frequentes variações das condições de rádio reportadas pelo UE.

Na Tabela 11 mostra-se o valor médio de RTT das sessões, que para este conjunto de testes deu o valor de 220,92 ms. O rácio da latência sentida no *downlink* com valor total do RTT, é de 88,3%. Observando a Figura 22, colocamos este operador entre a linha roxa e a linha verde, respectivamente entre os 300 ms e os 200 ms de latência. Isto indica que em média, para uma rede que disponibilize um valor de perda de pacotes na ordem dos 10^{-6}, este operador conseguiria possibilitar velocidade de *downlink* entre os 12 e os 19 Mbps, isto num cenário em que os parâmetros de rádio se encontram optimizados, os elementos de rede não se deparam de modo algum limitados na sua capacidade, e por fim que todo o licenciamento necessário para disponibilizar as referidas velocidades, nesses mesmos elementos de rede, tanto de acesso como de *core*, se encontram activos. De notar que as redes móveis de banda larga, 3G/WCDMA/HSPA, disponibilizam velocidades máximas de 84 Mbps. No 4G/LTE estas velocidades máximas podem atingir os 150 Mbps.

Aqui verifica-se que as latências sentidas nos dois sentidos são efectivamente distintas, o que aparentemente se poderia explicar pela diferença dos tamanhos dos pacotes existentes nos dois sentidos. Em *downlink* temos um ritmo elevado de pacotes com dimensão a rondar os 1460 *bytes* e em *uplink* pacotes de TCP ACK (*Acknowledge*) que rondam os 80 *bytes*. No entanto, no próximo capítulo, verifica-se que o comportamento em *uplink* de uma rede 3G/HSPA com as características de *Enhanced Uplink* de 2 ms como definidas nos 3GPP TS 25.306 [31], traz um ganho interessante ao tráfego de *uplink*.

Tabela 10 – Valores Gerais dos Testes *Download* Cidade A<->A

Latência [ms] DL CidadeA<->CidadeA	Origem	Destino	Média	Mínimo
E2E	Webserver	UE	197,72	14,00
Segmento 1	Webserver	Gn-SGSN	0,89	0,22
Segmento 2	Gn-SGSN	IuU	0,26	0,09
Segmento 3	IuU	UE	196,58	13,43
Componente do *Core*	Webserver	IuU	1,15	0,34

Tabela 11 – Relação da Latência de *Download* com o RTT nos Testes *Download* Cidade A<->A

CidadeA<->CidadeA	Origem	Destino	Média
Latência DL	Webserver	UE	197,72
RTT	Webserver	Webserver	220,92
Peso [Latência DL / RTT]			88,3%

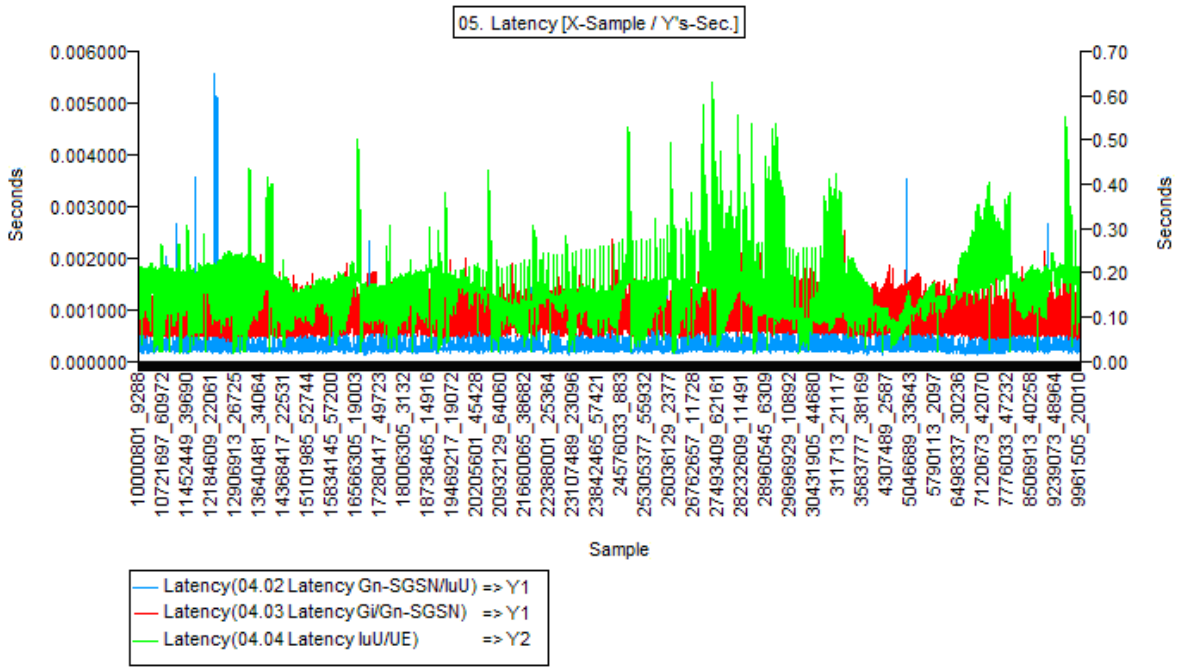

Figura 36 – Evolução da Latência DL por Amostra nos 3 segmentos de rede (Exemplo de uma Repetição) - Cidade A<->A

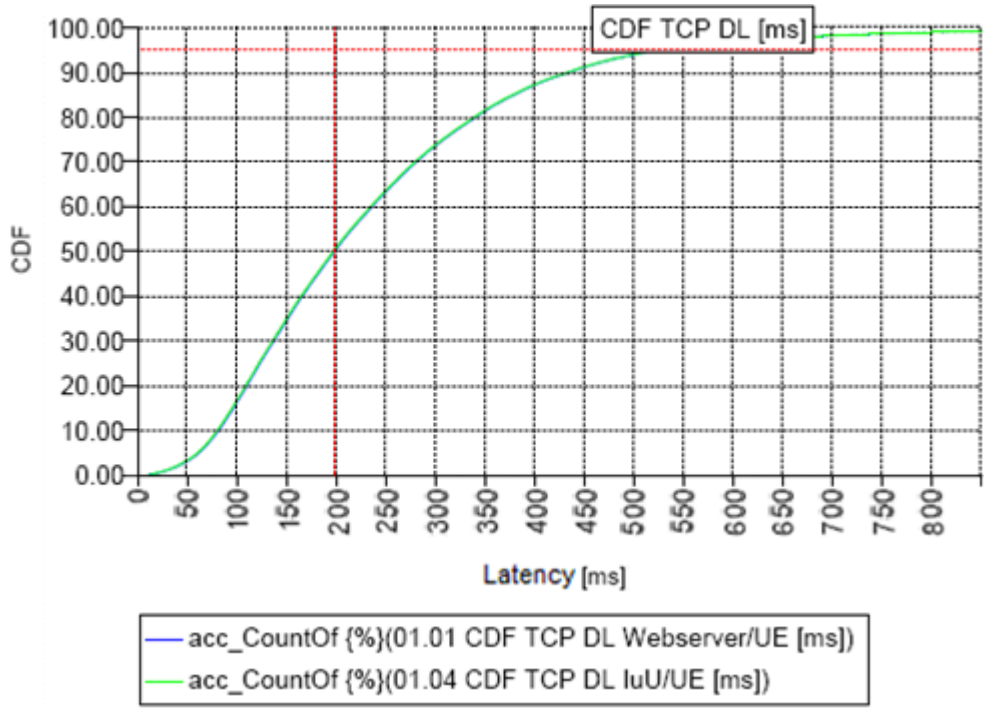

Figura 37 – CDF TCP DL - E2E (azul) e no Segmento 3 (verde) - Cidade A<->A

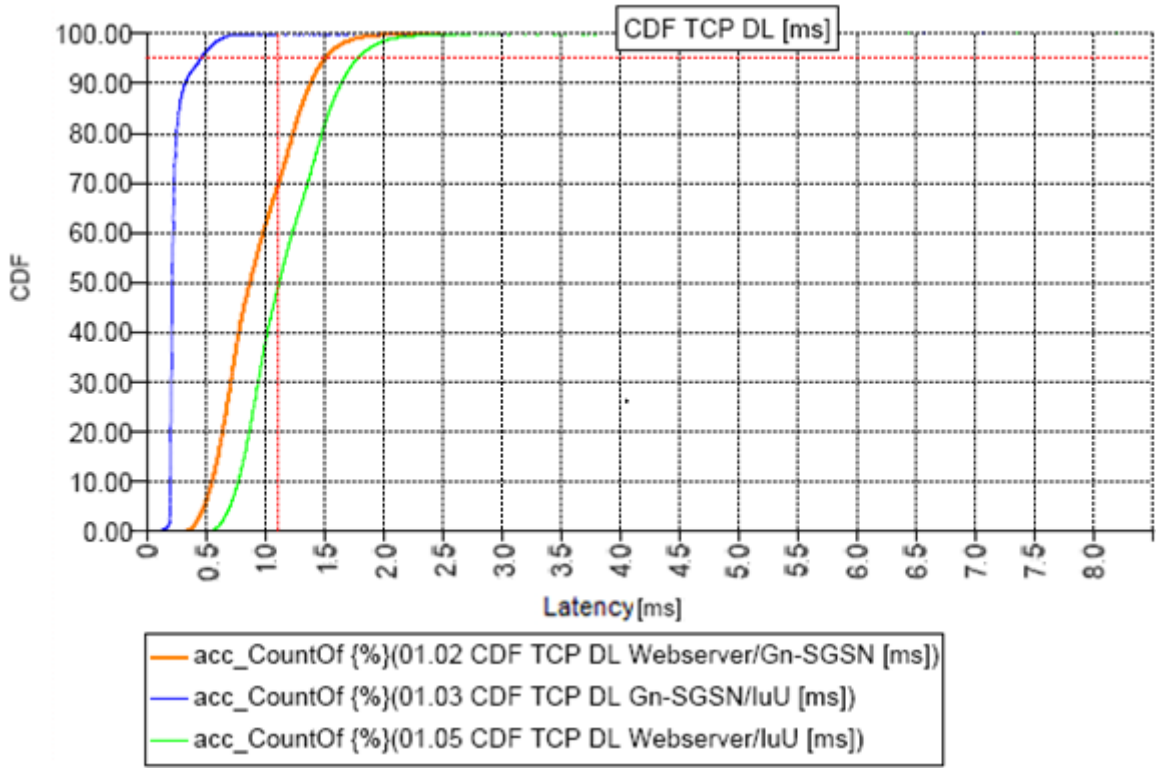

Figura 38 – CDF TCP DL – Componente de Core (verde) e no Segmento 1 (vermelho) e 2 (azul) - Cidade A<->A

7.1.2 Testes de *Upload* – SGSN e GGSN Co-localizados

Neste teste tentamos aferir os valores de latência, mas agora em *uplink* (UL), isto é, a enviar um ficheiro desde do *Laptop* ao Servidor localizado na rede de serviços do operador. Desta forma poderíamos sentir como toda a rede se comportava, com tráfego no qual o fluxo de pacotes de maiores dimensões era efectuado no sentido contrário ao que se efectuou no capítulo anterior.

O que mais salta à vista é o valor médio da perspectiva E2E de 51,96 ms e, como seria de esperar devido à sua percentagem de peso nessa ligação E2E, o comportamento do acesso rádio, representado na Tabela 12 pelo Segmento número 3. Neste último verifica-se um valor de 51,14 ms em média, o que é quase 4 vezes menor o que se obteve no mesmo troço em *downlink*. Estes valores de latência em *uplink* são possíveis devido às características de *Enhanced Uplink* de 2 ms existentes nesta rede móvel.

Os segmentos pertencentes ao *core* tiveram um melhor comportamento, mas como em média os seus valores já eram bastante baixos, continuam a não ter um contributo muito relevante no valor final da latência E2E, no entanto maior do que no teste em *downlink*, chegando agora aos 1,6%.

De uma visão RTT temos que o seu valor foi de 68,46 ms, fazendo com que o peso da latência em *uplink* tenha diminuído para os 75,90%. Olhando para a Figura 22, colocamos este troço do operador entre a linha cor-de-laranja e a linha azul escura, respectivamente entre os 100 ms e os 50 ms de latência. Isto indica que em média, para uma rede que disponibilize um valor de perda de pacotes na ordem dos 10^{-6}, este operador conseguiria possibilitar entre os 40 e os 70 Mbps, nas mesmas condições já indicadas no capítulo anterior. No entanto, as velocidades máximas permitidas numa rede 3G/HSPA em *uplink* são os 16 Mbps.

O valor do percentil 95 verificado neste teste na perspectiva E2E situou-se perto dos 140 ms, que é um valor muito mais interessante do que o conseguido no teste anterior. Este valor fica a aproximadamente 63% do valor médio obtido, o que significa a dispersão neste sentido é muito menor do que no teste anterior.

A dispersão no segmento do *core* continua a ser muito menor, principalmente no Segmento 2. No Segmento 1 existe uma dispersão um pouco mais longa, principalmente quando se atravessa o ponto médio e se vai no sentido do percentil 95.

Tabela 12 – Valores Gerais dos Testes *Upload* Cidade A<->A

Latência [ms] UL CidadeA<->CidadeA	Origem	Destino	Média	Mínimo
E2E	**UE**	**Webserver**	51,96	10,91
Segmento 1	**Gn-SGSN**	**Webserver**	0,61	0,16
Segmento 2	**IuU**	**Gn-SGSN**	0,21	0,07
Segmento 3	**UE**	**IuU**	51,14	10,38
Componente do *Core*	**IuU**	**Webserver**	0,83	0,26

Tabela 13 – Relação da Latência de *Upload* com o RTT nos Testes *Upload* Cidade A<->A

CidadeA<->CidadeA	Origem	Destino	Média
Latência UL	**UE**	**Webserver**	51,96
RTT	**UE**	**UE**	68,46
Peso [Latência UL / RTT]			75,9%

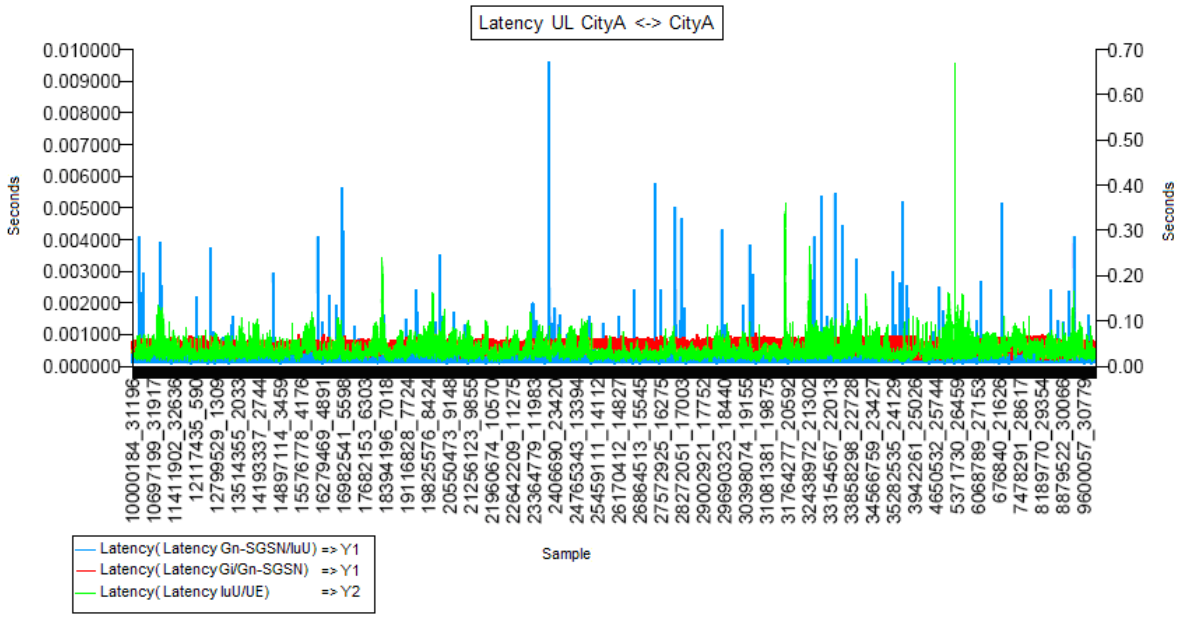

Figura 39 – Evolução da Latência UL por Amostra nos 3 segmentos de rede (Exemplo de uma Repetição) - Cidade A<->A

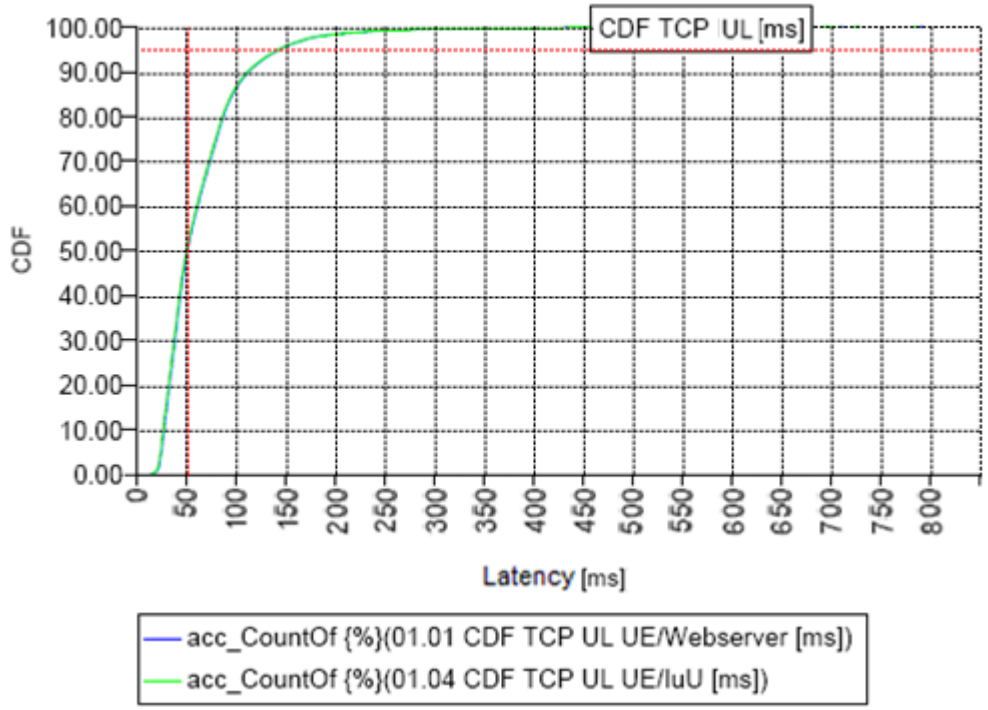

Figura 40 – CDF TCP UL- E2E (azul) e no Segmento 3 (verde) - Cidade A<->A

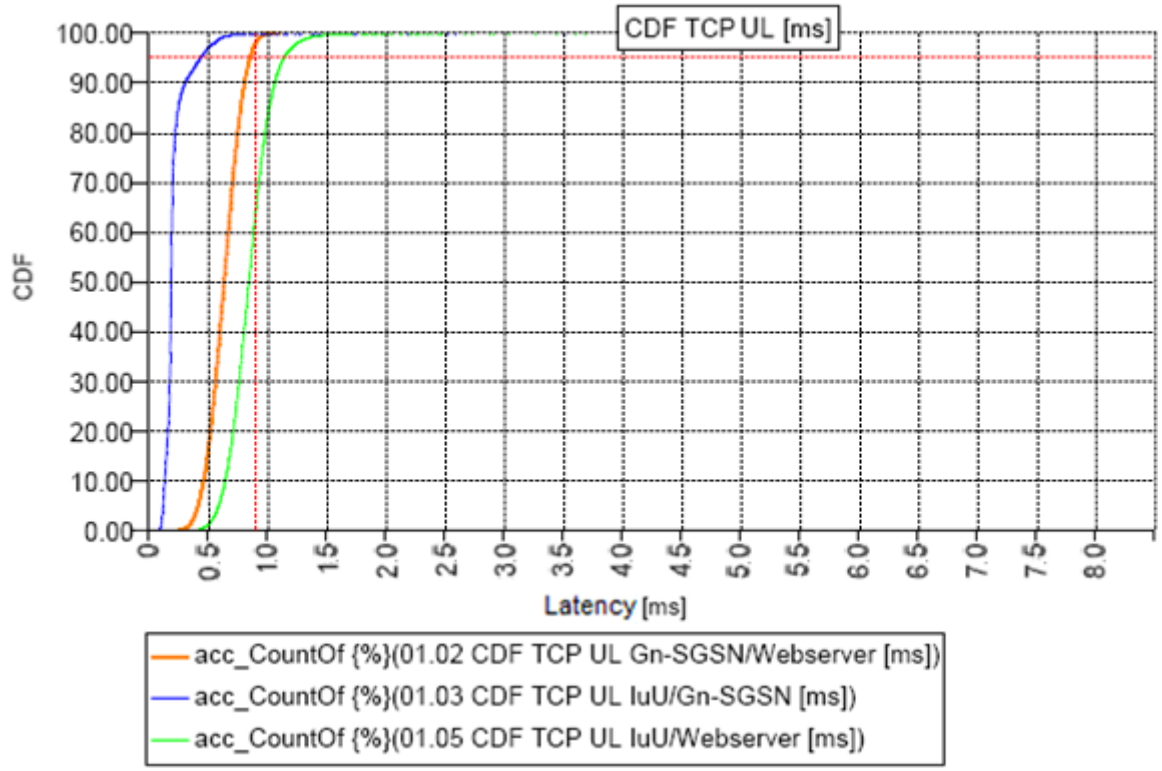

Figura 41 – CDF TCP UL – Componente de Core (verde) e no Segmento 1 (vermelho) e 2 (azul) - Cidade A<->A

7.2 Testes com Tráfego entre Cidades Distantes

7.2.1 Testes de Download – GGSN Distante

De forma a efectuar uma partilha dos recursos de rede na camada dos GGSN/PGW, é habitual os operadores fazerem um balanceamento aleatório, tipo *Round Robin*, dos PDPs activados em cada um dos GGSN's da rede. O racional deste teste é o de averiguar qual o verdadeiro impacto desta decisão de arquitectura de rede na latência introduzida no fluxo do subscritor. De notar que a distância entre as duas cidades era próxima dos 250 Km. Devido a esta diferença no fluxo do tráfego na rede, não se conseguiu capturar os dados na interface da Gn do SGSN, então não nos foi possível mostrar os valores dos Segmentos 1 e 2. No entanto, o propósito do teste é perfeitamente alcançado, pois continuamos a conseguir recolher a latência entre o *Webserver* e o IuU, o que representa a latência agregada de todo o segmento do *core*.

Analisando os valores obtidos, verificamos que a latência média determinada de uma forma E2E cresceu em média cerca de 23 ms, quando comparada com o mesmo cenário mas quando o tráfego fluía entre nós co-localizados. A latência média E2E verificada foi de 221,17 ms, com um valor mínimo de 17,82. A componente de *core* penaliza agora a latência E2E num único sentido cerca de 4 ms, passando o seu valor médio para os 5,40 ms e conseguindo um valor mínimo de 4,22 ms. Olhando para o valor que se obteve do peso da latência em *downlink* para o valor RTT, podemos aferir que este componente da rede terá um impacto no RTT de 6,31 ms em média, o que comparado com o valor determinado no teste indicado no capítulo 877.1.1 dá um incremento de aproximadamente 5 ms. De notar também que o troço entre a interface do IuU e o UE incrementou para os 215,81 ms, o que comparado com o valor obtido no capítulo 7.1.1, aumentou 20 ms. É um curioso comportamento, que se deve a se ter aumentado o RTT global, logo a diminuição do *throughput* máximo obtido.

Na Figura 43 verifica-se que o valor do percentil 95 das amostras se situa nos 600 ms, o que também está alinhado com o acréscimo médio próximo dos 20 ms, já referido no início deste capítulo.

A penalidade de levar o tráfego para um GGSN distante está visível na Figura 44. Aqui é clara a impressão digital deste troço, onde comparada com a Figura 38, é evidente o desvio no início da linha do CDF para o milissegundo 4,40. O valor do percentil 95 é neste caso próximo dos 6 ms.

Tabela 14 – Valores Gerais dos Testes *Download* Cidade A<->B

Latência [ms] DL CidadeA<->CidadeB	Origem	Destino	Média	Mínimo
E2E	Webserver	UE	221,17	17,82
Segmento 1	Webserver	Gn-SGSN	-	-
Segmento 2	Gn-SGSN	IuU	-	-
Segmento 3	IuU	UE	215,81	13,14
Componente do *Core*	Webserver	IuU	5,40	4,22

Tabela 15 – Relação da Latência de *Download* com o RTT nos Testes *Download* Cidade A<->B

CidadeA<->CidadeB	Origem	Destino	Média
Latência DL	Webserver	UE	221,17
RTT	Webserver	Webserver	258,38
Peso [Latência DL / RTT]			85,6%

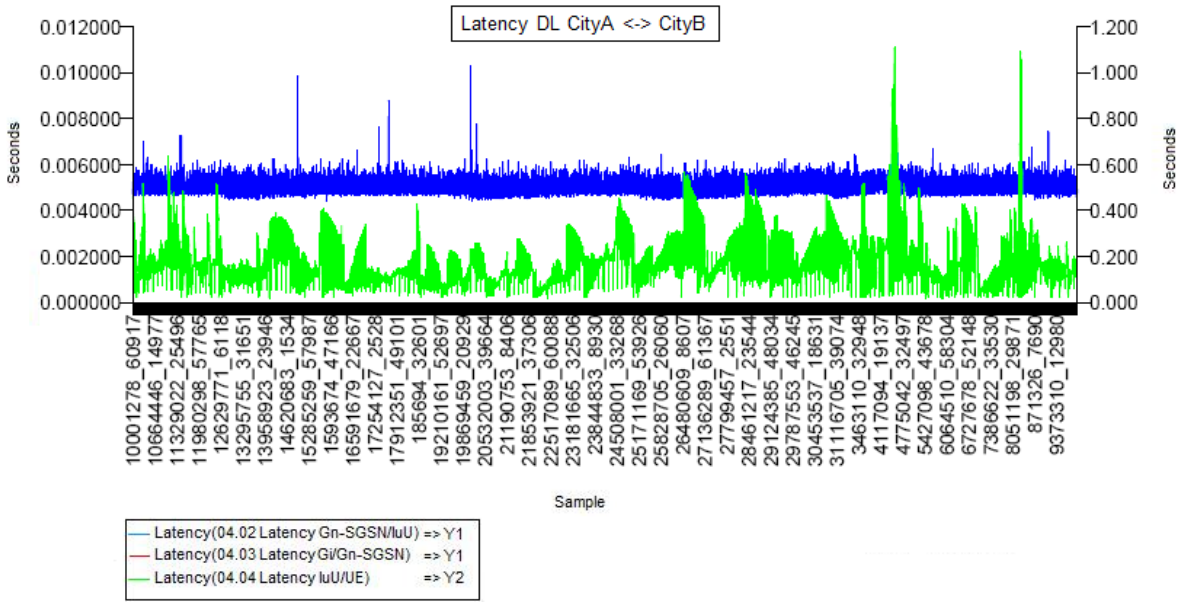

Figura 42 – Evolução da Latência DL por Amostra nos 3 segmentos de rede (Exemplo de uma Repetição) - Cidade A<->B

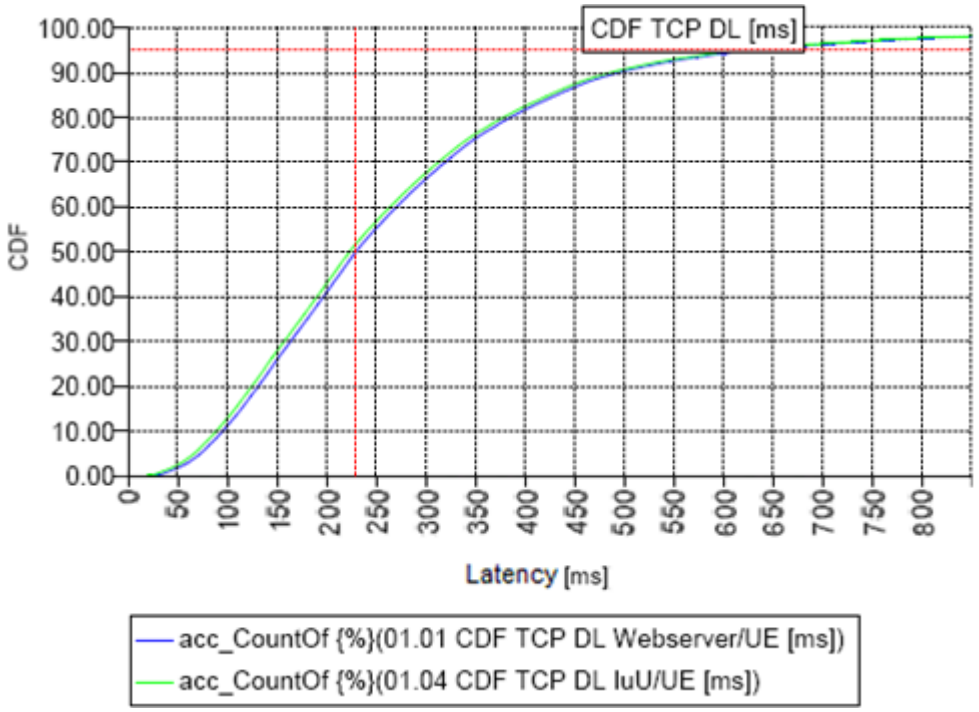

Figura 43 – CDF TCP DL - E2E (azul) e no Segmento 3 (verde) - Cidade A<->B

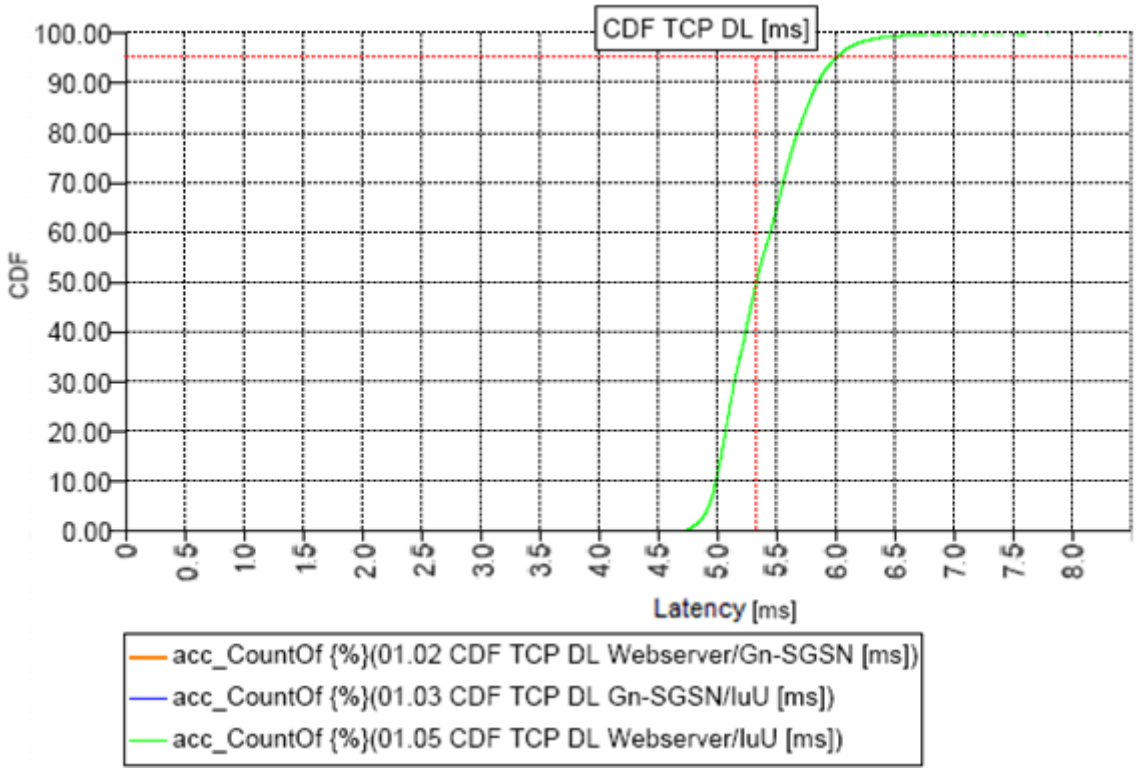

Figura 44 – CDF TCP DL – Componente de Core (verde) e no Segmento 1 (vermelho) e 2 (azul) - Cidade A<->B

7.2.2 Testes de *Upload* – GGSN Distante

Verificou-se neste teste, como já era expectável, que a latência de *uplink* E2E é consideravelmente melhor do que em *downlink*, como já tinha sido observada e explicada no capítulo 7.1.2. O valor obtido foi de 71,54 ms, o que representa aproximadamente 32% do valor obtido em *download*. O percentil 95 situa-se pelos 225 ms, cerca de 3 vezes superior ao valor médio obtido.

Os valores do *core* tiveram e continuam a ter um comportamento mais célere que o da componente de acesso, o primeiro continua a ter um valor médio a rondar os 5 ms o que representa um peso na latência em *uplink*, de aproximadamente de 7%. O percentil observado foi de sensivelmente 5,70 ms. Como era de esperar, estes troços continuam a ter uma previsibilidade mais estreita no tempo.

O *Round Trip Time* medido situa-se à volta dos 100 ms. Este valor proporciona um valor máximo de *Throughput* próximo dos 38 Mbps, se considerarmos uma rede com uma perda de pacotes perto dos 10^{-5}.

Tabela 16 – Valores Gerais dos Testes *Upload* Cidade A<->B

Latência [ms] UL CidadeA<->CidadeB	Origem	Destino	Média	Mínimo
E2E	**UE**	**Webserver**	71,54	15,30
Segmento 1	**Gn-SGSN**	**Webserver**	-	-
Segmento 2	**IuU**	**Gn-SGSN**	-	-
Segmento 3	**UE**	**IuU**	65,75	10,42
Componente do *Core*	**IuU**	**Webserver**	5,24	4,13

Tabela 17 – Relação da Latência de *Upload* com o RTT nos Testes *Download* Cidade A<->B

CidadeA<->CidadeB	Origin	Destination	Average
Latência DL	**UE**	**Webserver (IuU)**	71,54
RTT	**UE**	**UE**	100,20
Peso [Latência DL / RTT]			71.4%

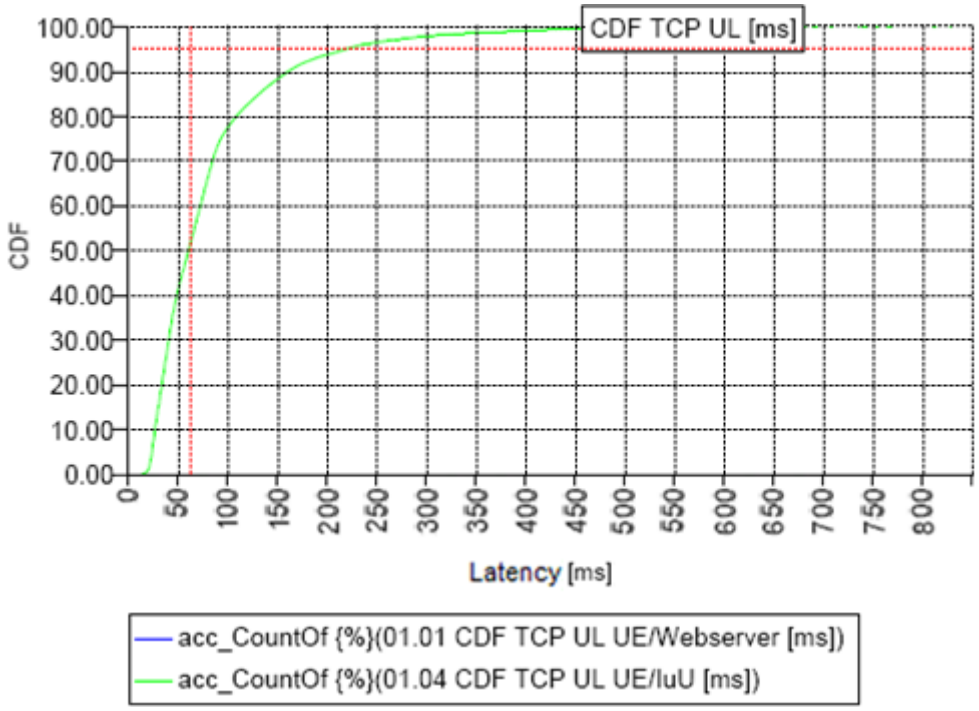

Figura 45 – CDF TCP UL - E2E (azul) e no Segmento 3 (verde) - Cidade A<->B

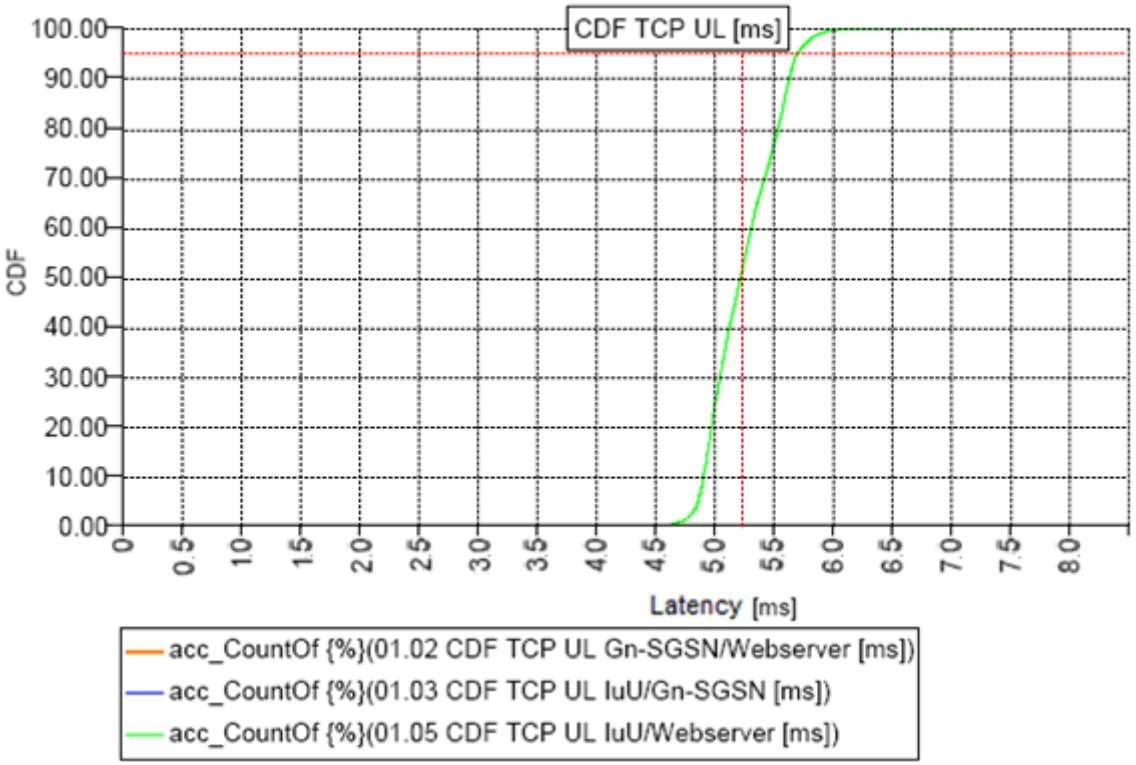

Figura 46 – CDF TCP UL – Componente de Core (verde) e no Segmento 1 (vermelho) e 2 (azul) - Cidade A<->B

7.3 Testes de Uplink 24 Horas – SGSN e GGSN Co-localizados

Com este teste, que é amplamente mais abrangente no tempo que os anteriores, conseguiu-se observar um comportamento que, com os métodos actuais disponíveis no mercado e indicados no capítulo 2.3.1, não seria possível observar.

7.3.1 Comportamentos e Evidências

As duas figuras seguintes indicam a perspectiva de uma sonda colocada na posição de um subscritor, num determinado lugar da rede móvel. A Figura 47 representa, a verde, o valor médio horário de latência da rede que a sonda iria reportar. Como se observa, ela reportaria uma latência de 125 a 140 ms no período nocturno, que se encontra compreendido da 1h00 até às 9h30 da manhã. No período mais exigente para a rede ao nível da carga dos nós, ela reportaria valores próximos dos 160 ms. Este período encontra-se definido das 21h00 até às 23h30. Portanto, consegue-se observar que a latência E2E tem alguma dependência da carga da rede.

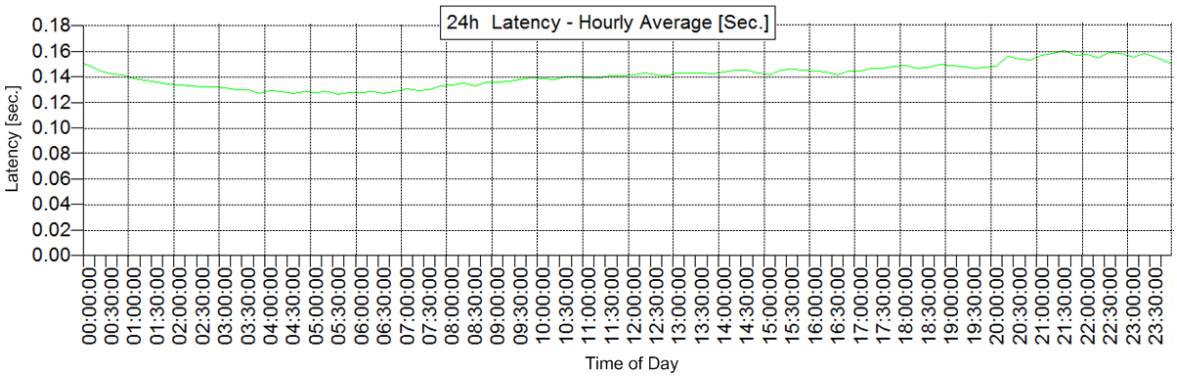

Figura 47 – 24 Horas – Média horária da Latência E2E da perspectiva de uma sonda

Se a sonda em questão reportasse os valores da latência por amostra, Figura 48, verificava-se que se observam alguns eventos esporádicos mas com alguma cadência, nos quais a latência observada salta para os 4 ou 5 segundos, com picos de 16 segundos, que daria um valor 25 a 100 vezes maior do que o valor médio reportado durante uma hora.

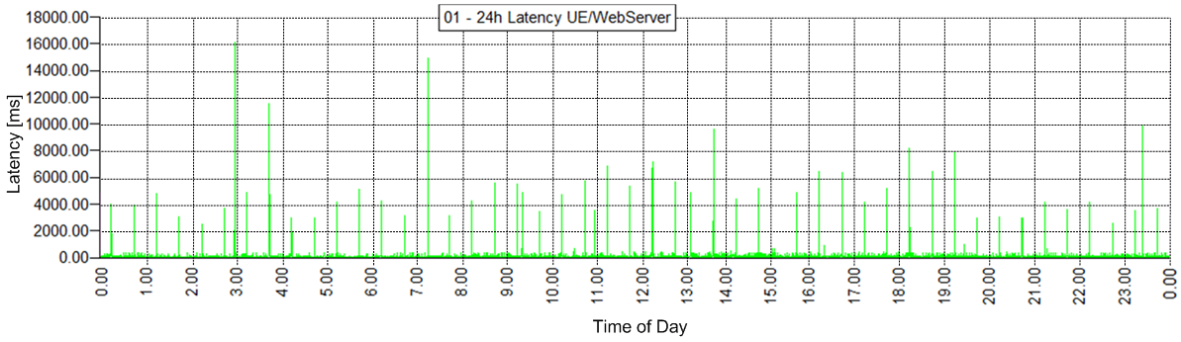

Figura 48 – 24 Horas – Latência E2E por amostra da perspectiva de uma sonda

Ampliando a escala vertical, Figura 49, observam-se de uma forma mais evidente dois factores. O primeiro é que existe uma mancha horizontal clara que vai dos 50 até aos 100 ms, o que evidencia que a latência média se encontra neste intervalo. O segundo factor é que existem com frequência amostras com valores substancialmente superiores ao valor médio da latência, que possivelmente advêm das condições de rádio.

Até este ponto dir-se-ia que este comportamento seria unicamente devido à parte de acesso, com especial incidência sobre a interface ar, pois é habitualmente considerado o "suspeito do costume".

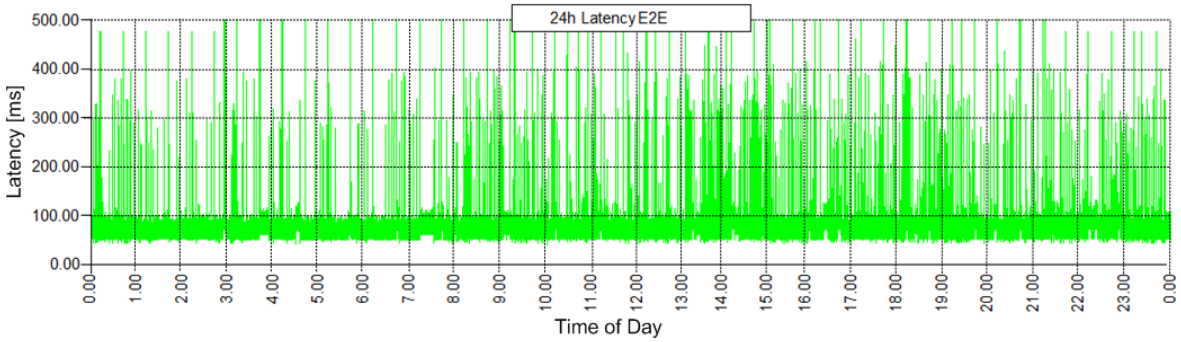

Figura 49 – 24 Horas – Escala ampliada da Latência E2E por amostra da perspectiva de uma sonda

Pegando agora nos 3 segmentos que constituem a troço E2E do fluxo seguido pelo tráfego do utilizador, vamos tentar validar se efectivamente o segmento 3, entre o UE e a interface IuU, é o responsável máximo do comportamento verificado, e que os dois segmentos de *core* não têm qualquer efeito relevante para acrescentar à latência E2E.

Verifica-se que a Figura 50 é à primeira vista muito parecia com a Figura 48, com o mesmo padrão já identificado na perspectiva E2E.

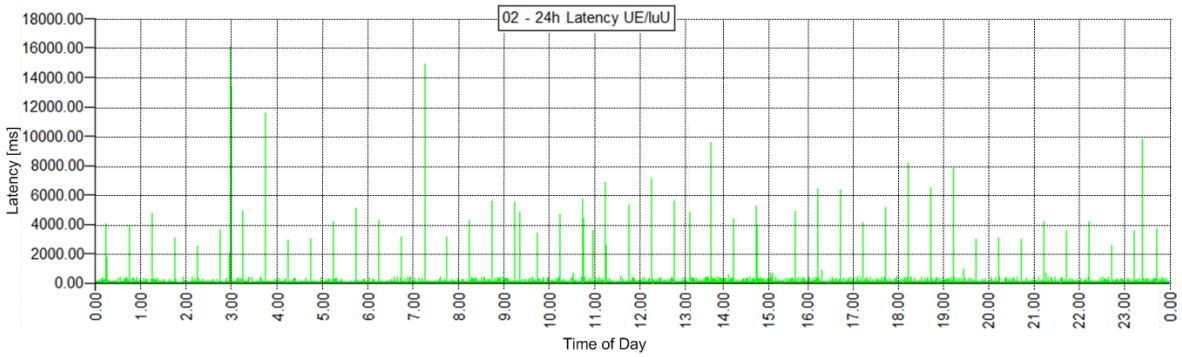

Figura 50 – 24 Horas – Latência no segmento 3, de acesso, entre o UE e o IuU, por amostra

Observa-se que o segmento 2, entre a interface IuU e a Gn, tem um comportamento muito aceitável, com latências médias abaixo do 1 ms, com uns picos acima dos 2 ms, chegando mesmo aos 16 ms, com alguma incidência entre as 9h00 e as 1h00 do dia seguinte. Não se observa um padrão concreto neste comportamento.

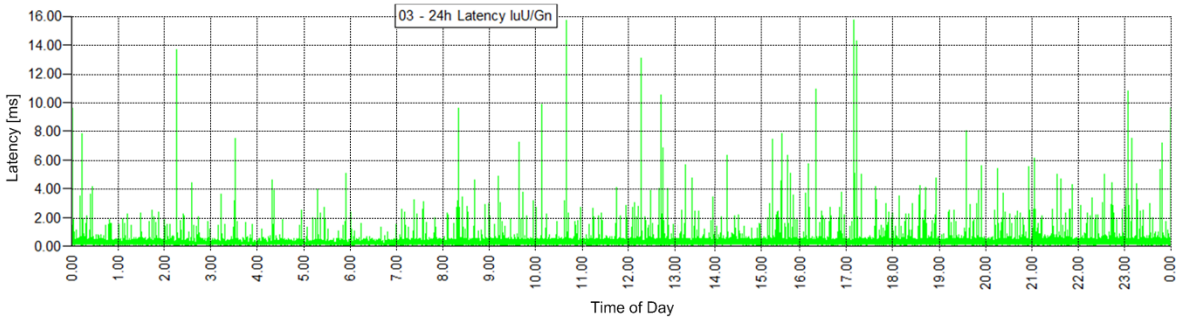

Figura 51 – 24 Horas – Latência no segmento 2, entre o IuU e a Gn, por amostra

Chegando ao segmento 1, entre a interface Gn e o *Webserver*, é clara uma relação directa entre o aumento da latência e a carga na rede. Este comportamento é observável na Figura 52 mas ainda mais na Figura 53, nas quais é evidente que das 2h00 até às 9h00 da manhã, a latência é estável, situando-se ente os 6 e o 9 ms. Quando os utilizadores despertam e começam a utilizar os seus dispositivos, pelas 9h00, a mancha verde tende para um intervalo situado entre os 20 e os 30 ms. Pelas 17h00 dá-se um novo incremento gradual até perto da 21h30, hora na qual se dá o pico de latência. Aqui o valor chega a situar-se acima dos 50 ms, o que representa um incremento de 10 vezes, quando comparada com o valor observado durante o período nocturno.

Este comportamento merece uma atenção por parte do operador pois não é de todo aceitável, principalmente quando se observa a Figura 27, a qual indica que a latência na tecnologia de acesso rádio LTE/4G irá evoluir para valores perto dos 5 a 10 ms. Este valor ficará então perto dos valores observados à noite no segmento 1, entre a Gn e o *Webserver*, mas

claramente distante dos valores observados na hora de pico deste mesmo interface. Assim o título “suspeito do costume” que é o acesso rádio, neste caso passaria para um dos segmentos do *core*.

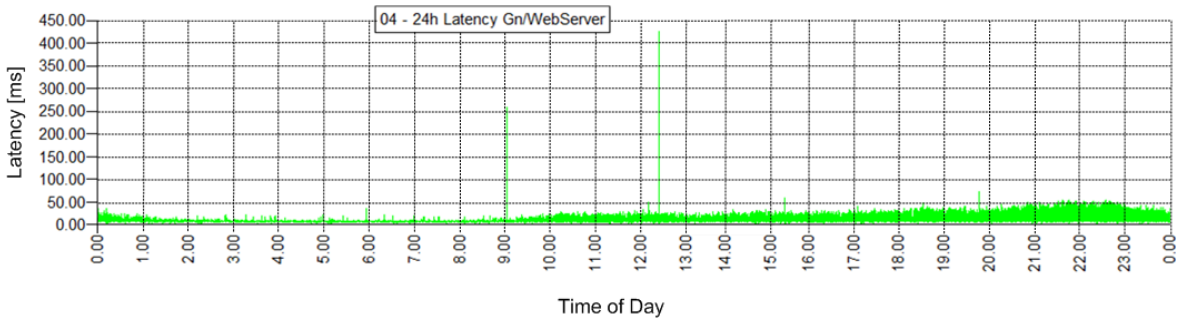

Figura 52 – 24 Horas – Latência no segmento 1, entre o Gn e o *Webserver*, por amostra

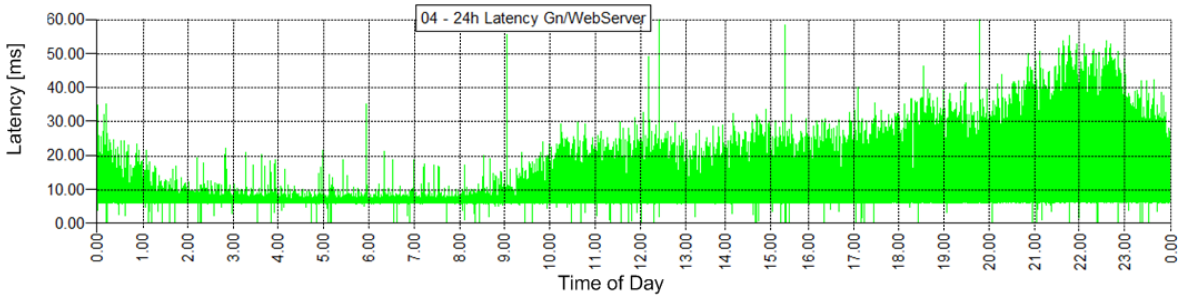

Figura 53 – 24 Horas – Escala ampliada da Latência no segmento 1, entre o Gn e *Webserver*, por amostra

Indo um pouco mais fundo no comportamento encontrado no segmento 1, verifica-se na Figura 54 que afinal, mesmo na hora de pico, existem valores de latência como existem no horário nocturno a rondarem os 7 ms. No entanto, existe um padrão com uma cadência de aproximadamente 2,5 minutos, no qual a latência dispara para os 40 ou 50 ms.

Este comportamento será devido a alguma rotina interna num elemento de rede no referido segmento, que é constituído por um GGSN e um PCEF, que tem impacto na capacidade desse elemento de rede processar o tráfego dos utilizadores de uma forma transparente, e independente da sua rotinas internas.

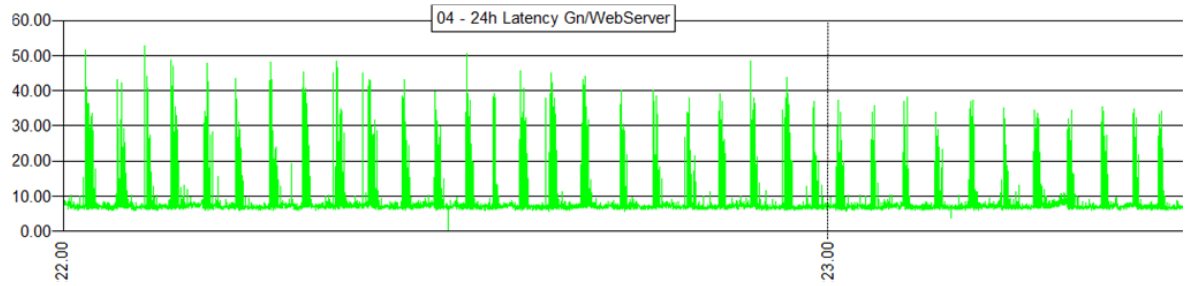

Figura 54 – 24 Horas – Escala ampliada da Latência no segmento 1, entre o Gn e *Webserver*, por amostra na hora de pico

Estes dois últimos comportamentos são uma evidência que a utilização de sondas unicamente numa perspectiva E2E não é suficiente para um operador móvel controlar e antecipar problemas na sua rede. A visão segmentada dessa mesma rede é essencial para se conseguirem observar eventos que em redes 3G/WCDMA/HSPA ainda podem ser tolerados, mas que em redes 4G/LTE poderão ter impactos que não serão de todo negligenciáveis na experiência de utilização.

7.3.2 Valores Medidos e *Benchmark*

O número de amostras recolhidas neste teste rondaram as 420.000, que permitiram caracterizar com bastante precisão a rede móvel do operador de onde foram recolhidas. De seguida vamos indicar os valores médios, mínimos, a mediana, os PDF's e CDF's por segmento de rede e também de uma forma E2E.

Iremos de seguida fazer um esforço de comparação, vulgo "*benchmark*", entre os dois operadores aferidos nos testes, tendo em consideração que a metodologia utilizada foi idêntica. Tendo em consideração a Tabela 12 no capítulo 7.1.2, verifica-se que esta rede tem um comportamento um pouco mais desvantajosa em termos de latência. Tem um comportamento em E2E com um acréscimo de 15 ms, no segmento 1 com uma penalidade de cerca de 4 ms. Os valores mínimos observados são também consideravelmente mais elevados tanto na perspectiva E2E, como no segmento 3 de acesso.

Tabela 18 – 24H - Latência de *Upload*

Latência ms	UE-IuU Segmento 3	IuU-Gn Segmento 2	Gn-Webserver Segmento 1	UE-Webserver E2E
Mediana	53	0	7	60
Média	58	0	7	66
Mínimo	30	0	1	36

Nas figuras seguintes podemos observar que o valor de incidência das latências na perspectiva E2E é próxima dos 60 ms. O segmento 3 tem o seu pico nos 52 ms. Por sua vez, o segmento 2 situa-se pelos 7 ms, e o segmento 1 situa-se pelos 0,2 ms.

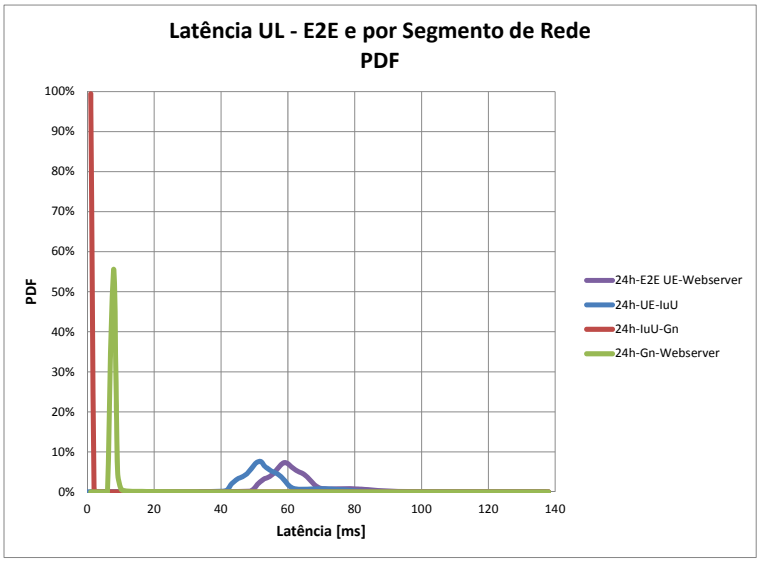

Figura 55 – 24 Horas – PDF dos vários segmentos de rede e UL

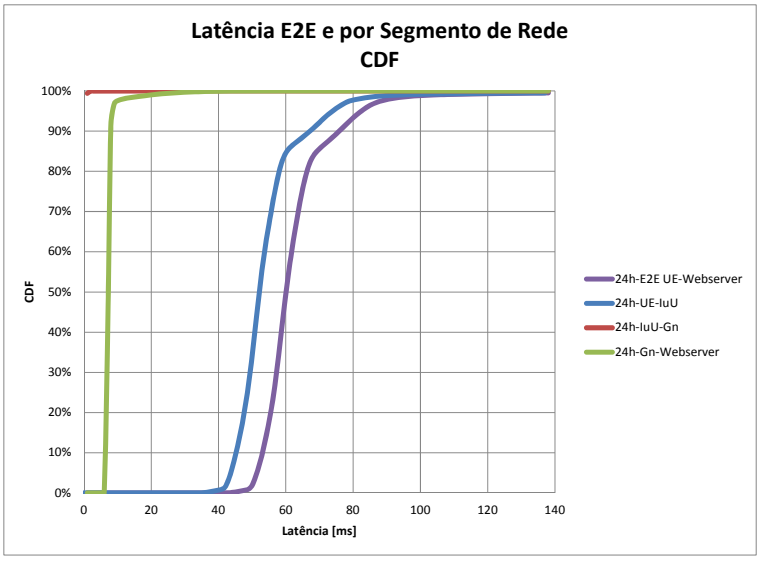

Figura 56 – 24 Horas – CDF dos vários segmentos de rede e E2E

7.4 Discussão

Ao se compararem os resultados obtidos nos vários testes e nos diversos segmentos, conclui-se que os valores de latência de *downlink* são substancialmente superiores aos valores de *uplink*. Isto deveu-se a dois factores, a existência da funcionalidade *Enhanced Uplink* com TTI's de 2 ms, e que o *downlink* na rede de acesso rádio no local onde se efectuaram os testes não se encontrar optimizada.

O segmento de rede que consome grande parte do valor de latência verificado entre as extremidades é o que fica compreendido entre a interface do UE e a interface Iu-PS, como seria de esperar pois é neste segmento do qual faz parte o interface ar. Por sua vez, o segmento que se situa entre a interface Iu-PS e a Gn é o que tem o melhor desempenho, logo seguida pelo segmento situado entre a interface Gn e o *Webserver*. Isto é válido em ambos os sentidos da transmissão.

Com a metodologia utilizada verificou-se que os dois sentidos do tráfego têm comportamentos completamente distintos em termos de latência. Verificou-se que o peso de tráfego no sentido do teste, onde se encontram os pacotes de dados com maior dimensão, é substancialmente maior do que no sentido oposto, no qual são enviadas as confirmações da correcta recepção dos dados recebidos. Isto reforça a teoria apresentada no capítulo 4.4, na qual se criticava a utilização de pedidos de ICMP como metodologia para se aferir com qualidade, o valor de *Round-Trip Time* de uma rede móvel.

A influência da distância entre o SGSN e o GGSN ficou também clara nos testes feitos. O impacto no desempenho da latência E2E, quando se recorre a uma distribuição de carga entre os diversos GGSN da rede, é algo que tem que estar claro na mente do departamento de engenharia do operador. De notar que actualmente, com a introdução do conceito de SGSN *in Pool*, esta mesma questão de levanta, agora numa posição mais a jusante na rede, pois a resiliência alcançada e desejada com esta arquitectura, não se coaduna com a optimização do valor de latência na rede, se não for acompanhada por metodologias que permitam optimizar o trajecto do fluxo do subscritor na rede. Esta atenção é principalmente relevante em situações nas quais o operador é posto à prova em testes de *benchmarks* de rede, nos quais um pequeno ganho no valor de latência pode fazer a diferença entre o segundo e o primeiro lugar.

Ficou também demonstrado que, se o operador conseguir ter esta metodologia implementada na sua rede, irá conseguir observar toda a cadeia de valor de latência de uma forma consistente e

conseguir retirar benefícios que ao dia de hoje não os consegue obter, que advêm do tipo de dados que se conseguem recolher.

O primeiro benefício do método apresentado, tendo ficado demonstrado nos quatro primeiros testes, está intimamente ligado com a experiência de utilização do subscritor. Com o tráfego capturado pela solução, permitiu aferir a qualidade que a rede disponibiliza a esse subscritor, pois como mostramos, transformando os valores de latência medidos nos dois sentidos, e sabendo o valor de perda de pacotes na rede, utilizando para tal os valores estatísticos dos diversos elementos de rede, recorrendo à Fórmula 1 consegue-se posicionar a sessão de dados do utilizador sobre o valor teórico de *throughput* máximo que a camada de TCP permite disponibilizar. Esta informação pode ainda ser mais trabalhada, de tal forma que pode ser utilizada para alertar áreas da rede que necessitem ser melhoradas, tanto a nível de parametrização como mesmo a expansões, recorrendo a aumentos de capacidade de transmissão ou novos elementos de rede.

Pode-se também alimentar um sistema de *Customer Experience Manager*, tanto para dar uma maior definição à qualidade da visibilidade entregue por esta ferramenta, como também melhorar o valor estimado do risco de *churn* para o operador em questão.

O segundo benefício, e que ficou evidente no teste de 24 horas, é que esta metodologia permite acompanhar a evolução da latência ao longo do dia em todos segmentos da rede de uma forma encadeada, e não de uma forma solta ou reduzindo-a à perspectiva E2E, como é habitual nas soluções existentes no mercado. Desta forma consegue-se captar tanto a tendência média das latências que se estão a verificar em cada um dos segmentos, como também observar comportamentos esporádicos ou padrões em algum desses segmentos de rede, que de uma perspectiva E2E estarão diluídos e que invariavelmente não se conseguiriam observar.

Estes valores podem também servir de referência para *benchmarks* internos, ou então para confrontar diferentes participações, numa operadora multinacional, de forma a aferir as diversas arquitecturas utilizadas e os diferentes fornecedores dos elementos de rede.

7.5 Conclusão

Neste capítulo descreveram-se os cenários criados e os testes utilizados para aferir a solução proposta nesta dissertação. Com a análise dos dados obtidos conseguiu-se concluir que os objectivos que nos tínhamos proposto foram alcançados e que a solução disponibiliza, de uma forma precisa, a quantificação da latência e que permite ao mesmo tempo visualizar a cadeia de valor dessa mesma latência de uma forma harmoniosa.

Foi avaliado o impacto de diferentes arquitecturas de rede e o seu impacto no valor global e parcelar da latência.

Evidenciou-se também qual a vantagem de se ter a solução proposta, a avaliar de uma forma permanente todos os interfaces da rede, revelando a sua utilidade na detecção de eventos e padrões que não seriam detectados de outra forma.

8. Conclusões e Linhas Futuras de Desenvolvimento

Nesta dissertação foi descrita e avaliada uma solução inovadora que consegue extrair informação das sessões de dados, e que permite quantificar de uma forma consistente um parâmetro de rede que tem especial impacto na qualidade de serviço, e que não é de todo habitual extrair com esta precisão e granularidade nas redes actuais. O referido parâmetro é a latência. A solução permite então avaliar a qualidade do serviço de banda larga e alertar para áreas da rede que necessitam de ser optimizadas.

A solução foi desenvolvida e testada em duas redes móveis de banda larga europeias que se encontram em serviço comercial. Os valores obtidos são boas referências para trabalhos futuros, e estes mostraram a importância das optimizações, tanto na parte de acesso de rádio como também na arquitectura de rede utilizada no *core*, para se minimizarem os valores de latência disponibilizados pela rede, e por consequência, maximizar a experiência do utilizador.

Os comportamentos verificados em alguns segmentos de rede são também relevantes, pois comprovaram que sem esta análise parcelar e encadeada da informação, proposta nesta dissertação, torna-se difícil captar alterações na latência que com as soluções existentes actualmente no mercado passariam completamente despercebidas, e que no entanto com a evolução actual nas redes móveis de banda larga, com a introdução do 4G/LTE no acesso rádio, podem ter um impacto catastrófico no desempenho global da rede.

O conceito da solução e da arquitectura proposta é versátil ao ponto de poder ser expandida sem quaisquer modificações, para as redes de banda larga móvel como as de 4G/LTE na parte de rádio, e EPC na parte de *core*. Isto é algo que ambicionamos fazer num futuro próximo. Pretendemos também expandir o alcance dos parâmetros de redes quantificados, passando a incluir o valor de perda de pacotes e *jitter* da sessão de dados. Desta forma, conseguiremos melhorar a qualidade dos valores obtidos que estão relacionados com a experiência de utilização, permitindo desta forma, não só colocar a sessão sobre as linhas da Figura 22, mas situar o ponto exacto.

Ainda em redes 4G/LTE, será interessante utilizar a metodologia proposta para aferir os diversos parâmetros de rede obtidos nos diversos QCI, que definem os padrões de qualidade impostos à rede. Assim daria confiança ao operador em colocar a sua voz sobre esta rede. Consideramos imperativo a um operador começar a ter essa visibilidade sobre os diferentes *bearers*, e que se encontram descritos na Tabela 9. Quando as chamadas de voz, ou seja de IMS,

começarem a ser transportadas nos seus *bearer*, isto significa uma grande mudança de paradigma para o operador móvel, pois ele irá começar a transportar numa rede desenhada para serviços de dados, um serviço que até à data era transportado por uma rede dedicada, e que tem um peso importantíssimo das suas receitas. Assim o operador tem que ter uma forma consistente de conseguir aferir em tempo real a qualidade disponibilizada pela sua rede, de forma a evitar a insatisfação dos clientes sobre o seu serviço de voz, e desta forma perder rendimentos.

Pretendemos também alargar o âmbito geográfico da captura dos dados, pois actualmente o equipamento que se utilizou nessa captura já suporta sincronização por GPS ou por IEEE 1588 (PTPv2), o que permite garantir a precisão entre os diversos dispositivos. Assim conseguia-se disponibilizar arquitecturas de rede mais complexas e mais enriquecedoras.

Referências

[1]. Vodafone Annual Report Accounts 2011, http://www.vodafone.com/content/dam/vodafone/investors/annual_reports/annual_report_accounts_2011.pdf. Vodafone Março 2011.

[2]. **C. Hedelin**, Superior network performance - a key differentiator for Network Service Providers. Ericsson Novembro 2012. http://www.ericsson.com/res/investors/docs/2012/investors-day/network-performance.pdf

[3]. **J. Gillet**, Wireless Intelligence: High-cost smartphone subsidies". Wireless Intelligence, Abril 2011.

[4]. **D. Parsons**, The Fixed/Mobile Broadband Battle: Is It Time for "Smart Broadband"?. IBSG, Novembro 2009.

[5]. **S. Téral**, 3G HSPA+ LTE Optimization: A Cook's Tour of HSPA+ in CEE. Infonetics Research, Inc., Abril 2012.

[6]. **ANACOM**, "Estudo de aferição da qualidade do serviço de acesso à internet banda larga". ANACOM, Abril 2009.

[7]. White Paper, "Transparent Network Performance Verification for LTE Rollouts". Ericsson, Setembro 2012.

[8]. Market Overview, "Data Dashborad", https://gsmaintelligence.com/markets/2798/dashboard/, GSM Intelligence, Q4 2012.

[9]. White Paper, "More than 50 Billion Connected Devices". Ericsson, Fevereiro 2011.

[10]. **G. Ragoonanan**, "Telecoms software professional services: Worldwide Forecast 2012–2016". Analysis Mason, Novembro 2012.

[11]. **P. Kelly**, **A. Rao**, "Service assurance systems: worldwide market shares 2012". Analysis Mason, Maio 2013.

[12]. IETF RFC 5357, Outubro 2008

[13]. Vodafone Group Plc - Preliminary Results - For the year ended 31 March 2012

[14]. White Paper, "The future of WCDMA/HSPA". Ericsson, Fevereiro 2013.

[15]. Packet Core Portfolio, http://www.ericsson.com/ourportfolio/products/packet-core, Ericsson, Junho 2013.

[16]. 3GPP TS 23.207, "End-to-end Quality of Service (QoS) concept and architecture", Dezembro 2008.

[17]. 3GPP TS 23.107, "Quality of Service (QoS) concept and architecture", Novembro 2011.

[18]. **P. Bates**, After the UK 4G auction, which UK mobile operator has the most valuable spectrum portfolio?. Analysys Mason, Março 2013.

[19]. **W. Lehr**, White Paper, Mobile Broadband and Implications for Broadband Competition and Adoption, Massachusetts Institute of Technology, 2010.

[20]. **M. Mathis**, **J. Semke**, **J. Mahdavi** e **T. Ott**, The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm. Computer Communication Review, Julho 1997.

[21]. **T. Szigeti** e **C. Hattingh**, "End-to-end QoS Network Design". Cisco Press, Setembro 2012.

[22]. 3GPP TS 25.331, “Radio Resource Control (RRC) protocol specification”, Novembro 2011.

[23]. 3GPP TS 25.322, "Radio Link Control (RLC) protocol specification", Junho 2010.

[24]. **S. Dixit** e **R. Prasad**, Wireless IP and Building the Mobile Internet. Artech House, 2003.

[25]. **J. Fabini**, **L. Wallentin** e **P. Reichl**, The Importance of Being Really Random: Methodological Aspects of IP-Layer 2G and 3G Network Delay Assessment. IEEE ICC, Junho 2009.

[26]. **M. Laner**, **P. Svoboda**, **P. Romirer-Maierhofer**, **N. Nikaein**, **F. Ricciato** e **M. Rupp**, A Comparison Between One-way Delays in Operating HSPA and LTE Networks. *In, 8th International Workshop on Wireless Network Measurements, 14-18 May 201.*

[27]. **T. Blajić**, **D. Nogulić**, **M. Družijanić**, Latency Improvements in 3G Long Term Evolution.

[28]. 3GPP TS 23.060 “General Packet Radio Service (GPRS)”, Dezembro 2010.

[29]. **P. Kelly**, Service assurance systems: worldwide forecast 2011–2015. Analysys Mason, Setembro 2011.

[30]. 3GPP TS 23.203, “Policy and charging control architecture”, Junho 2012.

[31]. 3GPP TS 25.306, "UE Radio Access capabilities", Junho 2011.